Pilot Contamination Mitigation for Massive MIMO System Using Novel Channel Estimation Techniques

Al-Salihi, Hayder Qahtan Kshash

Awarding institution:
King's College London

Pilot Contamination Mitigation for Massive MIMO System Using Novel Channel Estimation Techniques

by

Hayder AL-Salihi

Submitted to the Department of Informatics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

King's College London, University of London

May 2018

Author .....
Department of Informatics
May 16, 2018

Certified by .....
Dr Mohammad Reza Nakhai
Senior lecturer
Thesis Supervisor

Pilot Contamination Mitigation for Massive MIMO System Using Novel Channel Estimation Techniques

by

Hayder AL-Salihi

Submitted to the Department of Informatics
on May 16, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Massive multiple input multiple output (MIMO) is a promising technique to achieve the targets of the fifth generation of mobile communications. However, the pilot contamination problem creates a limitation to the potential benefits of massive MIMO systems. To mitigate the pilot contamination, in this thesis. novel channel estimation schemes are proposed and analyzed. First, an efficient channel estimation approaches based on Bayesian learning namely Bayesian compressed sensing (BCS) that rely on prior knowledge of statistical information about the channel sparsity is proposed for massive MIMO systems. Further enhancement has been proposed to the proposed technique of BCS through the principle of thresholding. Also, the multi-task BCS is also proposed to exploit the common sparsity distribution of the system channel. Furthermore, the Cramer Rao bound (CRB) has been derived as a reference line. Second, a novel channel estimation for massive MIMO systems using sparse Bayesian learning (SBL) is proposed. In the proposed technique, the sparsity of each channel coefficient is controlled by its own hyperparameter and the hyperparameters of its immediate neighbours. The mean square error (MSE) analytical expression for the proposed technique is derived. Based on that MSE expression, a pilot design criteria is proposed to design the optimal pilot to improve the estimation accuracy of the proposed algorithm.

Next, the optimal pilot for massive MIMO system has been investigated. The optimal pilots are designed by minimising the MSE of the minimum mean square error (MMSE) using the semidefinite programming (SDP) optimisation approach. Then, the conventional channel estimation is considered for Massive MIMO in a correlated Rician fading and correlated Nakagami-m fading channel models. Our analysis reveals that by increasing the line-of-sight (LOS) component the pilot contamination can be eliminated.

Finally, the discrete Fourier transform (DFT) based channel estimation is

proposed for massive MIMO, our simulation results show the effectiveness of the DFT channel estimation techniques for reducing the pilot contamination in comparison with the conventional based channel estimation.

Thesis Supervisor: Dr Mohammad Reza Nakhai
Title: Senior lecturer

This thesis is dedicated to my parents, wife and daughters

Acknowledgements

First, I thank Allah Almighty for giving me the strength and the means to complete my doctoral studies. I would like express my deep sincere thanks for my supervisor Dr Mohammad Reza Nakhai for his assistance and insight into my research. Next, I would like to thank Dr. Fatin Said and Dr Tuan Le for their helpful discussion and comments. I would also like to thank the research student at CTR. It is impossible to name them all. However, I like to give a special thanks to Ehsan, Sumayyah, Omar, Waleed, Nasreen, Christoforos, Firuz, Xinruo, Zhenfeng, Mohammed, Xun, Xi, Sunila and Syed. I wish to thank my parents, my brothers and sister and my little family for their morale support.

Finally, I would like to acknowledge the financial support of the higher committee for educational development (HCED) in Iraq.

List of Abbreviations

5G	Fifth Generation
AWGN	Additive White Gaussian Noise
BCRB	Bayesian CRB
BCS	Bayesian Compressed Sensing
BI-AMP	Bilinear Approximate Message Passing
CDF	Cumulative Distribution Function
CFR	Channel Frequency Response
CIR	Channel Impulse Response
CR	Compression Ration
CRB	Cramer Rao Bound
CS	Compressed Sensing
CSI	Channel State Information
DFT	Discrete Fourier Transform
DMSTA	DFT-based Most Significant Taps Approach
EM	Expectation Maximization
FDD	Frequency Division Duplex
FIM	Fisher Information Matrix
IDFT	Inverse DFT
LMI	Linear Matrix Inequality
LOS	Line of Sight

LS	Least Square
LTE-A	Long Term Evaluation-Advanced
MAP	Maximum a Posterior
MIMO	Multiple Input Multiple Output
MMSE	Minimum Mean Square Error
MRC	Maximal ratio combining
MSE	Mean Square Error
MST	Most Significant Taps
MT-BCS	Multi Task-Bayesian Compressed Sensing
MU-MIMO	Multi-User MIMO
NLoS	Non Line of Sight
OFDM	Orthogonal Frequency Division Multiplexing
OMP	Orthogonal Message Passing
RIP	Restricted Isometry Property
RLS	Regularized Least Square
SBL	Sparse Bayesian Learning
SDP	Semidefinite Programming
SISO	Single Input Single Output
SVD	Singular Value Decomposition
TDD	Time Division Duplex
WR	Whitening Rotation

Symbols

a	scaler a
$\mathbf{a}$	vector $\mathbf{a}$
$\mathbf{A}$	matrix $\mathbf{A}$
$\|\|\mathbf{A}\|\|_F$	Frobenius norm of $\mathbf{A}$
$\|\|\mathbf{A}\|\|^2$	Spectral norm of $\mathbf{A}$
$\mathbf{A}^*$	complex conjugate of $\mathbf{A}$
$\mathbf{A}^T$	transpose of $\mathbf{A}$
$\mathbf{A}^H$	complex conjugate transpose of $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of $\mathbf{A}$
$tr(\mathbf{A})$	trace of $\mathbf{A}$
det	is the determinant
$A_{i,j}$	denotes the (i,j)th element of
$E\{.\}$	denotes the expectation of random variables within the brackets.
$diag(x)$	denotes a diagonal matrix
$\mathcal{CN}(\mu, \sigma)$	complex scalar Gaussian random distribution with mean μ and variance σ^2
$\circledast$	denotes convolution operation
$\sum$	denotes summation operation
$\mathbb{C}^{n \times m}$	set of $n-$by$-m$ complex matrices
$\underset{\{\mathbf{Z}\}}{\text{minimize}}$	minimizes all the elements of $\mathbf{Z}$

s.t.	subject to
e^x	the natural exponential function
$\log(x)$	natural logarithm

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

The main activity of recent research has identified that the major targets for the next generation of mobile communications, the so-called fifth generation (5G) of mobile communications, are to achieve 1000 times the system capacity and 10 times the spectral efficiency, energy efficiency and data rate, and 25 times the average cell throughput [1], [2], as shown in Fig. 1-1. Therefore, it is necessary to introduce new technologies to meet the demand explosive requirement for 5G. From a high-level perspective, there is a promising technology that enables reaching higher fifth generation targets, called massive multiple input multiple output (MIMO). A massive MIMO can be defined as a system using a large number of antennae at the base station; accordingly, a significant beamforming can be achieved and the system capacity can serve a large number of users, as shown in Fig. 1-2 [3], [4].

The major challenge that limits the massive MIMO potential features is the acquisition of precise channel state information (CSI) at the base station. In general, based on the operating duplex mode, the acquisition of CSI approaches can be classified according to the following categories as time division duplex

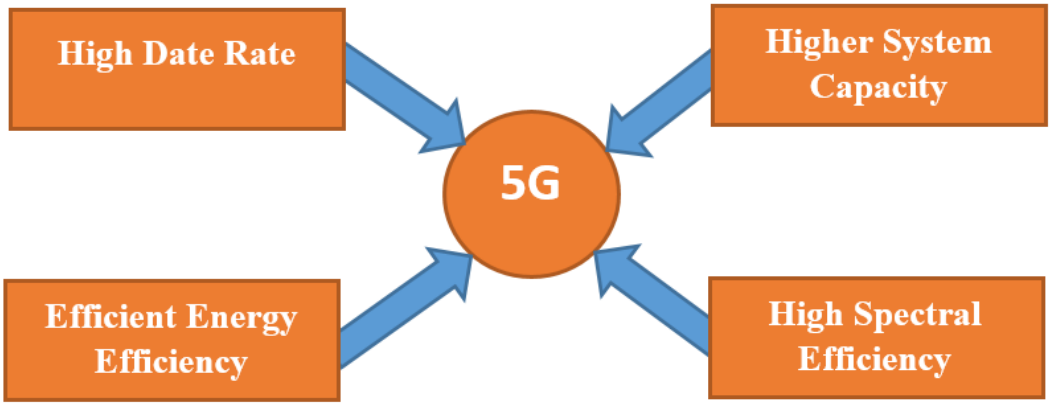

Figure 1-1: 5G Requirements.

(TDD) mode and frequency division duplex (FDD) mode. In FDD systems, the antennae at the base station send orthogonal pilots to the mobile stations and the channel will be estimated by the mobile station. The estimated channel will be then fed back to the base station. Hence, the number of orthogonal pilots is proportional to the number of antennae, which makes FDD impractical when employing massive MIMO, as the antenna array at the base station becomes very large. However, in TDD mode, each mobile station is assigned an orthogonal pilot that will be sent to the base station, and the base station will estimate the channel using an appropriate estimation algorithm. To attain the optimum channel estimation accuracy, perfect orthogonal pilots allocation to the users are required, unfortunately, this requirement is a challenge to achieve, because the number of pilots has to be proportional to the number of users in the system [5-8]. Failure to meet this requirement leads to the creation of a spatially correlated intercell interference, known as pilot contamination, which reduces the estimation performance and the spectral efficiency [5-8].

However, the pilot contamination problem could be reduced by decreasing the number of pilots. Hence, the development of efficient channel estimation

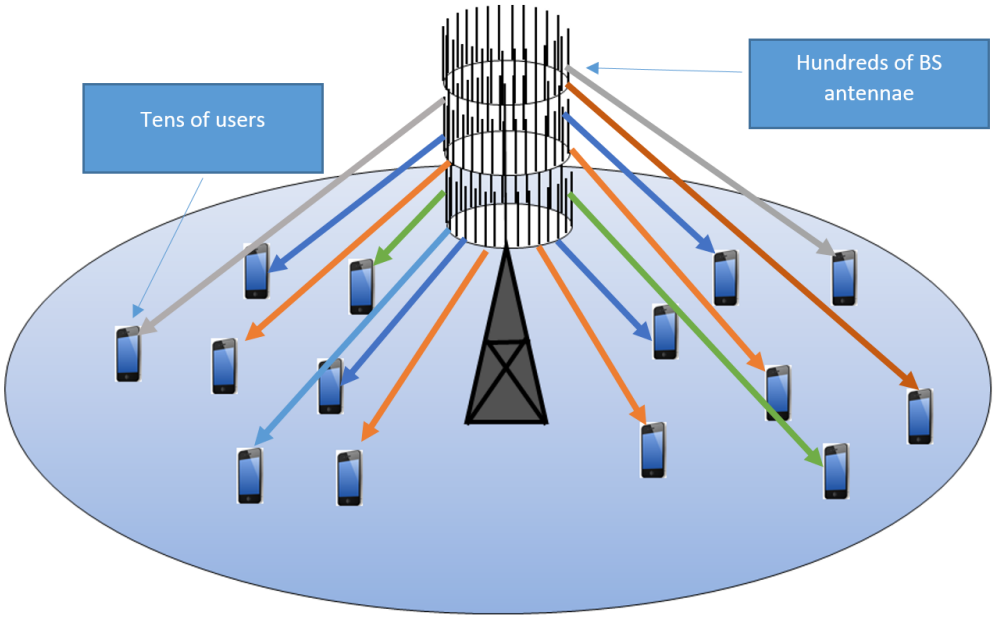

Figure 1-2: Illustration of massive MIMO system.

techniques that require a fewer number of pilots is a challenge that should be thoroughly addressed.

1.2 Previous Research

To tackle the pilot contamination problem in massive MIMO systems, several techniques have been proposed. the author in [9] attempted to address the pilot contamination problem by creating more orthogonal pilots by reducing the frequency reuse factor, so as results of that the adjacent cells operate on different frequencies. Thus, users in different cells have orthogonal pilots. However, a smaller frequency reuse factor means less bandwidth for each cell, which will reduce the system capacity [10]. The authors [11]-[15] proposed using precoding techniques via multi-cell cooperation. In this approach, the base stations are connected to a hub to exchange information via optical fibres. Each BS will upload data to a network hub. Then, the hub will encode the data through precod-

ing techniques and disperse them to the corresponding BSs. Each base station then conducts its own precoding to eliminate pilot contamination. This approach addresses the pilot contamination problem. However, information exchange between the base stations will add more complexity and overheads. The authors in [16] proposed a time shifted communication protocol. In this approach, the cells are divided into three groups, and different groups employ shifted communication schedules. When users of an arbitrary cell transmit pilot sequences, all the contiguous cells are in the phase of downlink data transmission or BS processing. As a result, pilot contamination can be mitigated to a great extent. However, a significant problem with this method is the highly restricted pilot length, leading to a smaller number of users per cell.

On the other attempts, channel estimation approaches are alternative solutions that can attain great improvements in contrast to multi-cell cooperation, frequency reuse deduction or protocol-based methods.

The authors in [17], [18] and [19] applied classical estimation algorithms,i.e., least square (LS) and minimum mean square error (MMSE) for massive MIMO systems. The LS based estimator is simple to implement, while the MMSE estimator yields much better estimation than the LS, although the MMSE imposes significantly higher complexity than LS. However, both techniques suffer from pilot contamination. In addition, the MMSE requires knowledge of the channel statistics. The authors in [24] -[30] proposed the superimposed channel estimation approach for massive MIMO systems by adding a training (pilot) signal with low power to the data signal at the transmitter. The superimposed signal is then utilised at the receiver for channel estimation. However, a proportion of the power allocated to the training signal is wasted, which has a negative impact on energy efficiency. In addition, this scheme causes inter-cell interference known as cross contamination. Motivated by the recent attention to compressed sensing (CS) techniques, channel estimation based on CS techniques has been investigated for massive MIMO systems in [31] - [32]. CS can recover unknown signals

using only a small number of measurements significantly fewer samples than is possible when using the conventional Nyquist rate, by exploiting the sparse nature of signals (that is, only a small number of components in a signal vector are non-zero). Thus, CS allows for accurate system parameter estimation with fewer pilots, thereby addressing the pilot contamination problem and improving the bandwidth efficiency [33]-[34]. However, classical CS algorithms require prior knowledge of channel sparsity, which is usually unknown in practical scenarios. In addition, to apply CS algorithms, the sampling matrix must satisfy the restricted isometry property (RIP) to guarantee reliable estimators. Such a condition cannot be verified easily because the condition are computationally demanding [33] - [34]. The authors of [35]-[36] proposed the blind estimation scheme to eliminate the pilot contamination problem completely. However, up to date, most of the proposed blind methods have major disadvantages, such as slow convergence speed, high computational complexity and poor performance when compared to pilot assisted methods.

Given the aforementioned background and motivation, the main objective of this thesis is to develop an efficient channel estimation scheme for massive MIMO system to eliminate/alleviate the pilot contamination problem without the need for the base station cooperation, frequency reuse reduction or relaying on protocol methods. Therefore, in this thesis the main focus is on channel estimation, which is the lowest complex approach to address the pilot contamination problem in massive MIMO system. Furthermore, the objective is to tame the shortages and limitations of the currently proposed estimation methods of CS, blind and superimposed channel estimation.

1.3 Contribution

Motivated by the potential features of massive MIMO, in this thesis, we proposed the following channel estimation algorithm.

- Bayesian Compressed Sensing

  In Chapter 3, we proposed an improved channel estimation scheme based on the theory of Bayesian CS (BCS) that relies on prior knowledge of statistical information about the channel sparsity to overcome the shortages of CS-based channel estimation. We also proposed enhancing the performance of the BCS-based estimator using the principle of thresholding to select the most significant taps to improve channel estimation accuracy. In addition, we exploited the common statistical sparsity distribution to enhance the estimation accuracy performance by proposing the multi task-BCS (MT-BCS) based estimator. Furthermore, to provide the benchmark for the minimum performance error of the BSC and MT-BCS, the Cramer Rao bound (CRB) has been drawn for BCS and it has been derived and drawn for MT-BCS. Our simulation results indicate that the proposed channel estimation methods provide improved estimation accuracy and can address the pilot contamination problem. The material of this contribution appears in the following publications:

  - Hayder AL-Salihi, Mohammad Reza Nakhai, "Bayesian Compressed Sensing-based Channel Estimation for Massive MIMO Systems", Proceeding of European Conference on Networks and Communications (EuCNC), Athens, Greece, 27-30 June 2016, EURAIP/EuCNC Best Student Paper Award.
  - Al-Salihi, H. Q. K. and Nakhai, M. R. "Efficient Bayesian Compressed Sensing-based Channel Estimation Techniques for Massive MIMO-OFDM Systems", 23 Feb 2017 In : EURASIP Journal on Wireless Communications and Networking (EURASIP JWCN).

- Sparse Bayesian Learning

  In Chapter 4, we proposed a novel channel estimation for massive MIMO systems, using sparse Bayesian learning (SBL) based on a pattern coupled

hierarchical Gaussian framework. In the proposed technique, the sparsity of each channel coefficient is controlled by its own hyperparameter and the hyperparameters of its immediate neighbours. The simulation results show that the channel coefficients can be estimated more efficiently in contrast to the conventional channel estimators in terms of channel estimation. Furthermore, we derive the mean square error (MSE) analytical expression for the proposed technique and based on that MSE expression, a pilot design criteria is proposed to design the optimal pilot to improve the estimation accuracy of the proposed algorithm using the Lagrange multiplier optimisation method. Results showed that we could reduce the MSE of the SBL estimator by employing the optimal pilot sequence. The material of this contribution appears in the following publications:

  - Hayder AL-Salihi; Mohammad Reza Nakhai and Tuan Anh Le "Enhanced Sparse Bayesian Learning-based Channel Estimation for Massive MIMO-OFDM Systems", Proceeding of European Conference on Networks and Communications (EuCNC), Oulu, Finland, 12-15 June 2017.
  - Hayder AL-Salihi; Mohammad Reza Nakhai and Tuan Anh Le "Enhanced Sparse Bayesian Learning-based Channel Estimation with Optimal pilot design for Massive MIMO-OFDM Systems", IET Communication, under review.

- Optimal Pilot Design

In Chapter 5, in order to estimate the CSI in the presence of pilot contamination accurately, we operated the conventional MMSE channel estimation process using an optimally designed pilot set to improve the performance of the proposed technique. The optimal pilots were designed to minimize the MSE under the total transmit power constraint based on optimisation

problem formulation. Thus, we could reduce the number of the employed pilots to mitigate the effect of pilot contamination. The material of this contribution appears in the following publication:

- Hayder Al-Salihi, Trinh Van Chien, Tuan Anh Le, Mohammad Reza Nakhai, "A Successive Optimization Approach to Pilot Design for Multi-Cell Massive MIMO Systems", IEEE communication letter, early access.

- Channel Model

  Thus far, most of the studies of massive MIMO systems have assumed the channel condition to be an independent and identically distributed (i.i.d.) Rayleigh fading. To evaluate massive MIMO in more realistic scenarios, we need models that capture important massive MIMO channel characteristics. Furthermore, the effect of correlation should be considered, as the majority of the previous studies assumed that the channels are independent. And in more realistic environments the antennae are not sufficiently separated and the propagation environment does not provide a sufficient amount of rich scattering [5]. Therefore, in Chapter 6, we investigated the multi-cell massive MIMO system in the correlated Rician fading and correlated Nakagami-m fading channel models. The material of this contribution appears in the following publications:

  - Al-Salihi, H. Q. K., Said, F., Nallanathan, A. and Wong, K. K "Estimation accuracy of multi-cell massive multiple-input multiple-output systems in correlated Rician fading channel", IET Electronics Letters, vol. 51, issue 22, pp 1830-1832, 29 October 2015.
  - Hayder AL-Salihi and Fatin Said "Performance Evaluation of Massive MIMO in Correlated Rician and Correlated Nakagami-m Fading", 9th International Conference on Next Generation Mobile Applications,

Services and Technologies (NGMAST), pp 222-227, Cambridge, 9-11 Sept. 2015.

- Discrete Fourier Transform (DFT)-Based Channel Estimation
  DFT-based channel estimation is a promising technique that is commonly used to improve the performance of the conventional channel estimation. The DFT estimation approach can significantly reduce the impact of the noise and the interference on the estimated channel coefficients. In Chapter 7, we proposed three modified DFT-based channel estimation techniques for massive MIMO systems. The proposed methods mitigate the pilot contamination significantly by exploiting the properties of DFT-based estimation through iterations and most significant (MST) approaches. Also, we proposed a whiting rotation (WR) semi-blind estimation for massive MIMO systems. Using such a scheme we can reduce the number of the required pilots to estimate the channel coefficient. The estimation accuracy of the DFT estimator can also be enhanced by combining DFT with the WR semi-blind estimation.

  - Hayder AL-Salihi and Mohammad Reza Nakhai "An Enhanced Whitening Rotation Semi-Blind Channel Estimation for Massive MIMO-OFDM in Correlated Fading Channel", ICT 16-18 May 2016, Thessaloniki, Greece.
  - Hayder AL-Salihi; Mohammad Reza Nakhai and Tuan Anh Le, "DFT-based Channel Estimation Techniques for Massive MIMO Systems", Intentional Conference on Telecommunication (ICT) 2018, accepted for publication.

1.4 Achievements

1. EuCNC 2016 student best paper award

2. EuCNC 2017 travel grant support.

Chapter 2

Background

2.1 MIMO Systems

The conventional wireless communication system employs a single transmit antenna at the transmitter and receiver sides that is known as a single-input single-output (SISO) systems. However, the wireless channel in such a system is severely attenuated when the channel in a deep fade and that will put limitation to the system capacity. This problems can be addressed by employing multiple antennae at the receiver, transmitter or both, whereby this scheme, the multiple antennae ensures an improved high-quality transmission by combating or exploiting fading. Thus, it is interesting to introduce the multiple antennae on both the receiver and transmitter sides and such a system is known as a multiple-input multiple-output or MIMO system. MIMO systems combat multipath by creating the spatial diversity technique, and those techniques that exploit multipath do so by performing spatial multiplexing [37]-[39].

For the spatial diversity technique, at any instant, the probability that all of these channels are influence by a deep fade is reduced significantly, thereby ensuring high system reliability, as can be shown in Fig. 2-1. Also, it can be seen in same figure that the MIMO systems can increase the throughput by exploiting

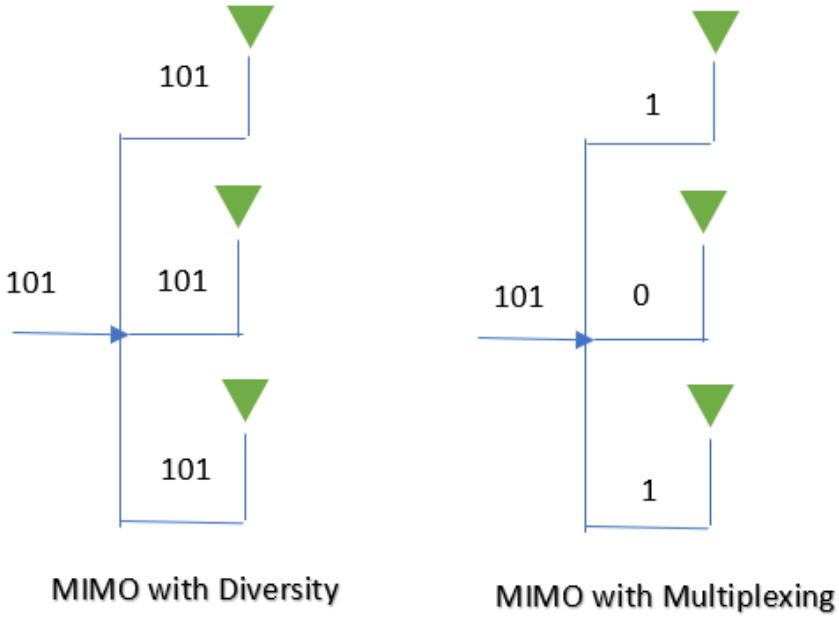

Figure 2-1: MIMO diversity system vsus MIMO multiplexing system.

the multipath by transmitting separate data streams over the transmit antennae and by separating those streams at the receiver using some form of spatial demultiplexing. Thus, a MIMO system offers the dual benefits of increased capacity due to spatial multiplexing and fading suppression due to receive/transmit diversity. These properties have greatly attracted the interest of researchers in the wireless community and in recent years there has been a flurry of activity in the MIMO area. In general, MIMO systems are classified into two categories: Point-to point MIMO and multi-user MIMO (MU-MIMO), as shown in Fig. 2-2. In the first category, the base station equipped with multiple antennae communicate with multiple antenna of a single user terminal. While in MU-MIMO, the base station is equipped with multiple antennae communicates with multiple users, and each user is with one or multiple antennae. Generally, MU-MIMO compared to single-antenna systems can provide better coverage, through beamforming that results in higher received signal power, and enhance the link reliability, through diversity

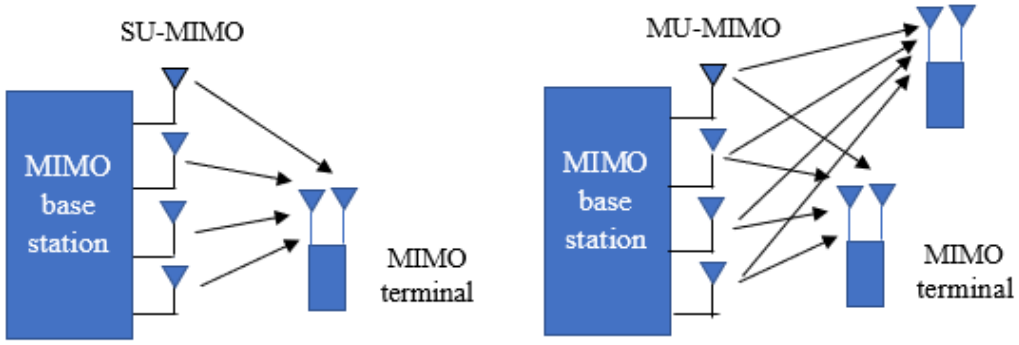

Figure 2-2: Single-user MIMO and multi-user MIMO.

schemes that combat fading effects in propagation channels and eventually reduce communication error probabilities, MU-MIMO can also provide higher capacity, through spatial multiplexing that transmits and receives several data streams in the same time-frequency resource [37]-[39].

2.2 Mobile Radio Channel

The communication medium between a transmitter and receiver (such as, mobile terminal and base station in a cellular system) is called a mobile radio channel. Radio wave propagation through these wireless channels is a complicated phenomenon characterized by various effects, such as noise, interference, and fading. Noise becomes the most common form of distortion in communication systems, resulting from the operating temperatures of the ohmic parts of the receivers, thereby limiting their sensitivity. The degradation in cellular networks can also result from intra-cell interference and inter-cell interference. Intra-cell interference is caused between frequency channels within the same cell due to adjacency of both frequencies and power leakage from one channel to an adjacent channel [40]. While in inter-cell interference is caused between a frequency channel in one cell and the same frequency channel used in another adjacent cell. On the other

hand, fading in the mobile radio channel results from the multipath propagation of the signal and can be divided into two broader categories, i.e. Large-scale fading and small-scale fading [41]-[42].

2.2.1 Large and Small Scale Fading

Large-Scale Fading

It is represented by the average signal power attenuation due to motion over large areas and is affected by prominent terrain contours (hills, forests, man-made structures, clumps of buildings, etc). The statistics of large-scale fading provide a way of computing an estimate of the average signal strength loss as a function of distance. This is described in terms of path loss and shadowing [40]-[41].

- Path Loss

  The decay of the mean signal power with distance from the transmitter is defined as path loss. It includes all of the possible elements of loss associated with interactions between the propagating wave and any object between the transmit and receive antennae. In free space, the mean signal power decreases with the square of the distance from the transmitter. In wireless channels, where often no direct LoS path exists between the transmitter and receiver, the signal power decreases with a power higher than two and is typically in the order of three to five [41]-[42].

- Shadowing

  It is also caused by the obstruction of the transmitted waves by hills, building, walls, trees, etc., resulting in some paths with increased loss, while others less obstructed reaching the receiver with increased signal strength. This varying signal strength exhibit a log-normal behaviour and generally modelled by log-normal distribution [41]-[42].

Small-Scale Fading

Commonly known as multipath fading, this refers to the drastic changes in signal amplitude and phase that can be experienced as a result of small changes in spatial separation between the receiver and the transmitter. It occurs as a result of reflection, scattering and diffraction of the transmitted electromagnetic wave at natural and man made objects. A multitude of waves arrive from many different directions with varying delays, amplitudes and phases, caused by the mobility of either the receiver, the moving objects or both, in the wireless channel. At the receiver, the superposition of these waves result in amplitude and phase variations of the composite received wave resulting in time-variant multipath propagation. The varying signal strength due to small-scale fading is highly sensitive even to small movements on the order of the wave length and may result in a totally different wave superposition [41]-[42].

2.2.2 Modelling of Wireless Multipath Channels

Statistical characterization of wireless fading channels has remaind a very important research area and considerable efforts have been devoted for the accurate modelling of such propagation environments. The results of these intense efforts provide a range of relatively simple and accurate statistical models for wireless channels which depend on the particular propagation environment and the underlying communication scenario. In this section, a review of Rayleigh, Rician and Nakagami-m fading models is presented [43].

Rayleigh Model

The Rayleigh distribution is frequently used to model multipath fading in NLoS environments between the transmitter and receiver. By letting the channel amplitude as α, the Rayleigh probability density function (PDF) can be given as

[44]

$$P_\alpha(\alpha) = \frac{2\alpha}{\Omega} \exp(-\frac{\alpha^2}{\Omega}), \tag{2.1}$$

where Ω is the average power of α

Rician Model

Also known as Nakagami-n, this model is commonly employed to characterize propagation environments consisting of one strong direct LoS path with many random indirect weaker components. When there is a line-of-site component between transmitter and receiver, the signal will be composed of large number of independent paths plus line-of-site components. The signal envelope is modelled by Rician probability density function distribution given by [44]

$$P_\alpha(\alpha) = \frac{2(1+n^2)e^{-n^2}\alpha}{K} \exp[-\frac{(1+n^2)\alpha^2}{K}] I_0(2n\alpha\sqrt{\frac{1+n^2}{K}})\alpha \leqslant 0, \tag{2.2}$$

where I_0 is the zeroth-order modified Bessel function of first kind and n is the Nakagami-n fading parameter ranging from 0 to ∞, which is also related to the Rician K factor by $K = n^2$. The parameter K represents the ratio of the power received in the LoS path to the total power received via indirect scattered paths [44].

Nakagami-m Model

This fading model, introduced by Nakagami in early 1940′s, has received considerable attention due to its great versatility in terms of flexibility and accuracy in providing a better match to various empirically obtained measurement data than Rayleigh or Rician distributions. It often gives the best fit to land, indoor-mobile multipath propagation as well as scintillating ionospheric radio links [44]. The

Nakagami-m PDF is in essence a central chi-square distribution given by [44]-[46]

$$P_{\alpha}(\alpha) = \frac{2}{\Gamma(m)}(\frac{m}{\Omega})^{m}\alpha^{2m-1}\exp(-\frac{m}{\Omega}\alpha^{2}), \alpha \leqslant 0, \tag{2.3}$$

where $\Gamma(.)$ is the Euler Gamma function and m is the Nakagami-m fading parameter ranging from 0 to 1 [44].

2.3 Massive MIMO Overview

Theoretically, for point-to-point MIMO and the MU-MIMO systems, the more antennae that those systems are equipped with, the better performance can be achieved in terms of reliability and the system capacity. The current fourth generation system of long term evaluation-advanced (LTE-A) employ up to eight antennae ports at the base station, one can ask, are the potential of MIMO being fully exploited? Thomas L. Marzetta published a paper in 2010 titled "Non-cooperative Cellular Wireless with Unlimited Numbers of Base Station Antennae" to investigate the potential of MIMO with a large number of antennae. The paper concluded that by employing an unlimited number of antennae at the base station the effects of receiver noise and fast fading are eliminated completely, and transmissions from terminals within the same cell do not interfere. However, transmission from terminals in other cells that use the same pilot sequence constitute a residual interference known as pilot contamination. In contrast to conventional MU-MIMO with up to eight antennae, the MIMO technology with a large number of antennae "massive MIMO", "very-large MIMO" or "large-scale MIMO" will increase the spectral and energy efficiency, system capacity and throughput [3], [4] and [6].

Considering the massive MIMO benefits, massive MIMO systems as a new technique suffers from some challenges i.e. hardware impairment, antenna Array implementation, pilot contaminations and others. However, the main objective

of this thesis is address the pilot contamination problem [4].

2.4 Pilot Contamination Problem

The major limiting factor in massive MIMO is the availability of accurate, instantaneous channel state information (CSI) that describes channel properties of a communication link at the base station. The CSI is typically acquired by transmitting predefined signals (pilot signals) and estimating the channel coefficients from the received signals by applying an appropriate estimation algorithm [2] and [30]. Channel estimation accuracy depends on having perfect orthogonal pilots allocated to the users; however due to shortages of the orthogonal signals, the same carrier frequency should be used in the neighbouring cells by following a specific reuse pattern. This leads to the creation of a spatially correlated inter-cell interference, known as pilot contamination, which reduces the estimation performance and spectral efficiency [2] and [30].

As elaborated in chapter 1, several methods have been proposed to address the pilot contamination problem i.e. the protocol-based method, precoding process approaches and channel estimation techniques [11]-[16]. In this thesis, we will try to address the pilot contamination problem using novel channel estimation methods.

2.5 Channel Estimation Philosophies

Channel estimation techniques can be classified into three categories i.e. training, blind and semi blind based channel estimation [47]. In this section, we introduce the pilot based estimation in subsection 2.4.1, followed by blind estimation in subsection 2.4.2 and the semi-blind estimation is discussed in subsection 2.4.3.

2.5.1 Pilot based Estimation

The conventional wireless communication systems in pilot based channel estimation employ a sequence of symbols that is known as a frame. Each frame comprised of pilot and data symbols. The pilot symbols follow a fixed set of symbols that is known at the receiver. So, at the receiver, one can estimate the channel by the observed received signal and known sequence of pilots. Then, the receiver can detect the transmitted data symbols. This system is referred as pilot-based channel estimation and considered as the common approach being employed in wireless communication system. This approach provides the advantageous of high quality estimation and low complexity. However, the major drawback of this approach is the pilot sequence carry no data so it will affect the overhead of the communication system and leads to bandwidth inefficiency [39] and [43].

2.5.2 Blind Estimation

Blind estimation is a scheme that does not employ predefined pilot and rely of the received data to estimate the channel by exploiting the statistical information of the data. Applying the blind method would eliminate the need for transmitting the predefined pilot sequence and the system would be totally eliminate the pilot contamination. However, this approach is suffered from high computationally complex and convergence problem [39] and [43].

2.5.3 Semi-Blind Estimation

As shown in the previous subsections, there is a tradeoff between the complexity and robustness versus the bandwidth efficiency and pilot contamination elimination depends on the employed estimation scheme whether pilot or blind estimation. So, is it possible to construct a scheme with a limited number of pilots whereby, it is possible to reduce the blind channel estimation shortages, and also

employ statistical information. This scheme is known as semi-blind estimator as it employs the pilot and blind information [39] and [43].

2.6 Probability Theory

Probability is defined as the fraction of time that a specific outcome occurs in an experiment which continues indefinitely. Where an experiment might be tossing of a coin or it could be receiving a noisy data that provide information about unknown radio channel in a digital communication system [49].

2.6.1 Tools of Probability Theory

This subsection provides the background on probability theory required to grasp most of the theoretical aspects.

Joint Probability

Joint probability is the probability of event Y occurring at the same time event X occurs. Let A and B events from an experiment, Then, the joint probability $P[A, B]$ of event A ocuuring at the same time with event B is defined as follows [49].

$$P[A, B] = P(A)P(B). \tag{2.4}$$

Conditional Probability

In most cases, knowledge about an event leads an additional information about the occurrence of another event. Let A and B events from an experiment, Then, the conditional probability of A given B $P[A/B]$ is defined as follows [49].

$$P[A/B] = \frac{P(A, B)}{P(B)}. \tag{2.5}$$

Independent Events

Statistical independence of two events means that the occurrence on one events has no influences on the occurrence of the other. Mathematically, two events A and B are statistically independent if and only if [49].

$$P(AB) = P(A)P(B). \tag{2.6}$$

A corollary of this definition is

$$P(AB) = P(A), \tag{2.7}$$

and

$$P(AB) = P(B). \tag{2.8}$$

Bayes' Theorem

More elaboration will be discussed in section 2.8.

2.7 Compressed Sensing

In signal processing, the Nyquist-Shannon sampling theorem states that analogue signals, images, videos can be perfectly recovered from from its samples if the samples are taken at a rate at least twice the bandwidth of the signal. However, the recently developed field of compressed sensing has successfully overcome the Nyquist-Shannon sampling theorem requirements. Compressed sensing techniques allow to sample a signal at a lower than Nyquist rate without any significant loss of information [50]. To review the main ideas of CS, Let $\mathbf{x} \in \mathbb{C}^{N\times 1}$ be signal vector with elements $x[n]$, $n = 1, 2, ..., N$, where x is a sparse vector, thus $\mathbf{x}$

is a vector with very few non-zero components and $\mathbf{\Psi} \in \mathbb{C}^{M \times N}$ as a measurement matrix with $M \leq N$. The observed signal is expressed as a vector

$$\mathbf{y} = \mathbf{\Psi}\mathbf{x} + \mathbf{n}, \tag{2.9}$$

where $\mathbf{y} \in \mathbb{C}^{M \times 1}$ and $\mathbf{n} \in \mathbb{C}^{M \times 1}$ is the additive noise in the system. The target of compressed sensing is to recover the signal $\mathbf{x}$ with no or insignificant loss of information using small number of N measurements. The best solution for this problem is to exploit recently developed CS techniques.

In general, the developed compressed sensing techniques can be classified into two categories. The non-Bayesian compressed sensing (BCS) that is the set of techniques that are based on convex optimization [6, 7, 8]. These techniques generally have excellent recovery performance and also have guaranteed performance bounds. High computational cost is one major drawback of these techniques, which makes it difficult to apply these techniques to large-scale problems. In addition these techniques are not practical as those technique assume that the knowledge of the signal sparsity [50].

Alternative techniques include BCS whereby we exploit prior statistical about the signal sparsity to tame CS deficiency. More elaboration regarding the Bayesian estimation philosophy will be presented in the next section.

2.8 Bayesian Estimation

In common literature, channel estimation methods are classified into parametric and Bayesian approaches. A standard parametric approach is the best linear estimator which is often referred to as least squares (LS) channel estimation. In contrast to parametric methods, the Bayesian approach treats the desired parameters as random variable with a-priori known statistics. Clearly, a priori PDF of the channel is assumed to be perfectly known at the receiver [51]-[52]. The esti-

mation of unknown parameters is the expectation of the posterior probabilistic distribution, where the posterior distribution is proportional to the prior probability and the likelihood of the unknown parameters based on the Bayes' rule, as can be shown in Fig. 2-3. For more clarifications about sparse Bayesian estimation, let consider the same target in the previous section of estimating the sparse signal $\mathbf{x}$ from observed signal $\mathbf{y}$ in (2.4) [51]-[52]. Thus the posterior distribution can be expressed based on Bayes' rule as follow

$$P(\mathbf{x}/\mathbf{y}) \propto P(\mathbf{y}/\mathbf{x})P(\mathbf{x}), \tag{2.10}$$

where $p(\mathbf{x}, \mathbf{y})$ is the likelihood function of $\mathbf{x}$ and $P(\mathbf{x})$ is the prior distribution of $\mathbf{x}$.

By applying the maximum a posterior (MAP) estimator which is the expectation of the posterior distribution, the estimated sparse vector can be expressed as

$$\hat{\mathbf{x}} = E[P(\mathbf{x}/\mathbf{y})]. \tag{2.11}$$

2.9 Bayesian Compressed Sensing

BCS provides an approach for solving the compressed sensing problem based on machine learning. In BCS, it is assumed that the unknown sparse vector follows a known statistical distribution. And the noise vector is sampled from some stochastic distribution (e.g. the multivariate Gaussian distribution). The main goal is now to estimate the parameters of the underlying distributions through the MAP. More precisely, to estimate the vector $\mathbf{x}$ in sections 2.7 and 2.8. It follows from Bayes rule that

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\text{argmax}}[P(\mathbf{x}/\mathbf{y})] = \underset{\mathbf{x}}{\text{argmax}}[P(\mathbf{y}/\mathbf{x})P(\mathbf{x})], \tag{2.12}$$

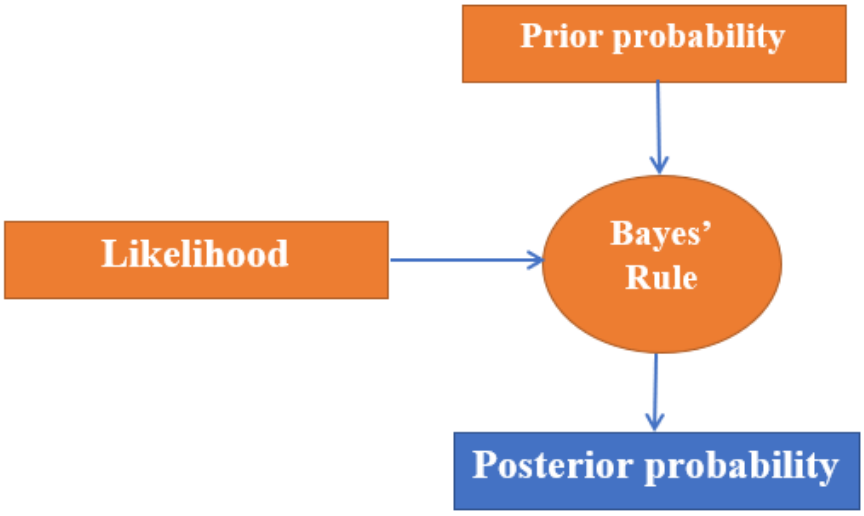

Figure 2-3: Bayesian rule

The distribution $P(\mathbf{x})$ models the noise process. The simplest noise model assumes that the measurement noise is white Gaussian of zero mean and variance $\sigma^2\mathbf{I}$, therefore we have

$$P(\mathbf{y}/\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-||\mathbf{\Psi x}-\mathbf{y}||_2^2/2\sigma^2} \tag{2.13}$$

Let assume that the a priori that the unknown sparse vector follows Gaussian distribution of Zero mean and ζ variance.

$$P(\mathbf{x}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\mathbf{x}^H\mathbf{x}\right)^2/2\zeta} \tag{2.14}$$

Now the conjugate of two Gaussian distribution is a Gaussian distribution with means and variance given by

$$\boldsymbol{\mu} = (\sigma^2)^{-1}\mathbf{\Sigma}(\mathbf{\Psi})^T\mathbf{y}, \tag{2.15}$$

$$\mathbf{\Sigma} = (\zeta \mathbf{I} + (\sigma^2)^{-1}(\mathbf{\Psi})^T \mathbf{\Psi})^{-1}, \tag{2.16}$$

Based on MAP estimation, the $\hat{\mathbf{x}}$ is the mean of the posterior distribution. Now, to calculate the values of σ^2 and ζ, the type II maximum likelihood procedure is employed where the conditional posterior distribution $P(\mathbf{y}/\mathbf{x})$ is first assembled algebraically. The maximum of the distribution is then found by taking partial derivatives of the distribution with respect to σ^2 and ζ in turn to obtain σ^2 and ζ to calculate the $\hat{\mathbf{x}}$ [53]-[54].

2.10 Convex Optimization

Convex optimisation is considered as the most widely research area in optimisation, in which the objective function and the constraint are convex. There are many advantageous of formulating the problem as a convex optimization problem. The most basic advantageous is that the problem can be solved more efficiently over other methods

Recent developments in convex programming extend the results and algorithms of linear programming for more complicated convex programs, e.g., conic programming. A conic programming is a linear programming with generalised inequalities. This section concisely reviews a standard conic program, i.e. semidefinite programming (SDP). Readers interested in convex optimisation and applications of convex optimisation in communications are referred to [55]-[57] for more details.

2.10.1 Semidefinite Programming

The standard form of a SDP is defined as

$$\begin{aligned} \min \quad & \mathbf{c}^H \mathbf{x} \\ \text{subject to} \quad & F(x) \succcurlyeq 0, \end{aligned} \tag{2.17}$$

where

$$F(x) \triangleq F_0 + \sum_{i=1}^{m} x_i F_i, \tag{2.18}$$

is called as the linear matrix inequality. For more calcifications and explanation about the SDP, lets give an example of how a convex optimization can be cast as SDP, Consider the problem

$$\begin{aligned} \min \quad & \frac{(\mathbf{c}^T \mathbf{x})^2}{\mathbf{d}^T \mathbf{x}} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} + \mathbf{b} \succcurlyeq 0, \end{aligned} \tag{2.19}$$

By introducing an auxiliary variable $\mathbf{S}$ that serves as an upper bound on the objective:

$$\begin{aligned} \min \quad & \mathbf{S} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} + \mathbf{b} \succcurlyeq 0, \\ & \frac{(\mathbf{c}^T \mathbf{x})^2}{\mathbf{d}^T \mathbf{x}} \preccurlyeq \mathbf{S}. \end{aligned} \tag{2.20}$$

By this formulation the convex objective in (2.9) is transferred to constraint in (2.10). The objective is a linear function of the variables $\mathbf{x}$ and $\mathbf{S}$. Now the constraints can be written as a linear matrix inequality in the variables $\mathbf{x}$ and $\mathbf{S}$ using Schur complements [55]-[57]

$$\begin{bmatrix} \mathbf{S} & \mathbf{c}^T \mathbf{x} \\ \mathbf{c}^T \mathbf{x} & \mathbf{d}^T \mathbf{x} \end{bmatrix} \succeq 0, \tag{2.21}$$

Thus, the convex problem in (2.9) has formulated as semidefinite program [55]-[57].

2.11 Conclusion

This chapter review the principles of massive MIMO systems. Then the pilot contamination problem is introduced. The methods of channel estimation are presented. Additionally, this chapter also provided an overview of mobile radio channel model. Then, a brief section about probability theory is presented. Thereafter, the Bayesian estimation and CS are introduced. Lastly, An approach to transfer the non-convex optimization problem using SDP approach is presented. It is to be hoped that the technical background provided in this chapter delivers the readers with the basic knowledge of the work areas to have better understanding on the proposed problems in the forthcoming chapters.

Chapter 3

Bayesian Compressed Sensing Channel Estimation Techniques

As stated in the previous chapters, in massive MIMO systems, the achievable estimation accuracy is limited in practice due to the problem of pilot contamination. It has recently been shown that compressed sensing (CS) techniques can address the pilot contamination problem. However, as elaborated in Chapter 1 that the CS-based channel estimation techniques suffers from practical shortages. To overcome these shortages, in this chapter, an efficient channel estimation approach is proposed for massive MIMO systems using Bayesian compressed sensing (BCS) based on prior knowledge of statistical information regarding channel sparsity. We have also proposed to enhance the performance of the BCS-based estimator through the principle of thresholding to select the most significant taps to improve the channel estimation accuracy. Furthermore, by utilising the common sparsity feature inherent in the massive MIMO system channel, we extend the proposed Bayesian algorithm to a multi-task (MT) version, so the developed MT-BCS can obtain better performance results than the single task version. To provide the benchmark for the minimum performance error of the BSC and MT-BCS, the Cramer Rao bound (CRB) has been drawn for BCS and it has been

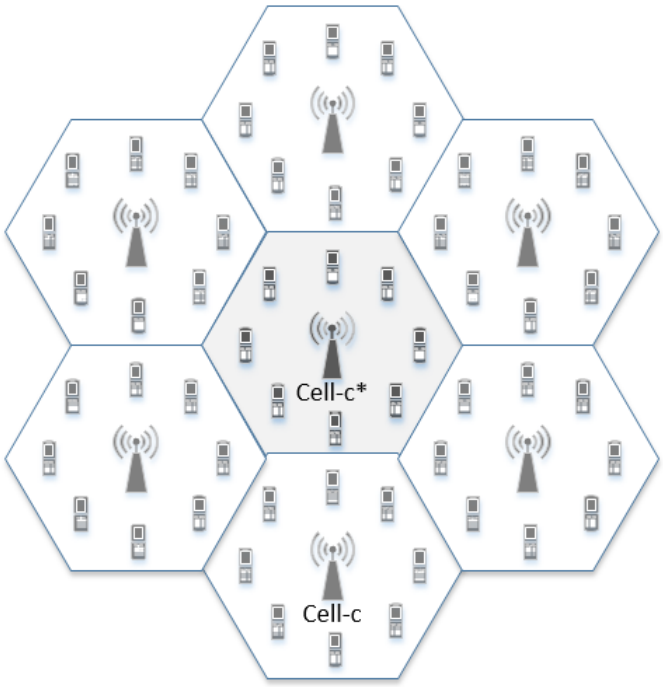

Figure 3-1: Illustration of the system model of a multi-cell multi-user massive MIMO.

derived and drawn for MT-BCS. Several computer simulation based experiments are performed to confirm that the proposed methods can reconstruct the original channel coefficient more effectively when compared to the conventional channel estimator. The remainder of this chapter is organized as follows: The multi-cell massive MIMO system model is presented in Section 3.1. The BSC-based and the MT-BSC based channel estimation details are reviewed in Section 3.2 and Section 3.3, respectively. In section 3.4, we provide the Cramer-Rao bound analysis. Section 3.5 presents the simulation results. Finally, the final conclusions are drawn in Section 3.6.

3.1 Massive MIMO System Model

We consider a time division duplexing (TDD) multi-cell massive MIMO system with C cells as shown in Fig. 3-1. Each cell comprises of M antennae at the

BS and N single antenna users. To improve the spectral efficiency, orthogonal frequency division multiplexing (OFDM) is adopted as in [58],[59].

At the beginning of the transmission, all mobile stations in all cells synchronously transmit OFDM pilot symbols to their serving base stations. Let the OFDM pilot symbol of user n in the c-th cell be denoted by $\mathbf{x}_c^n = [x_c^n[1]\ x_c^n[2] \cdots x_c^n[K]]^T$, where K is the number of subcarriers. The OFDM transmission partition the multipath channel between the user and each antenna of the BS into K parallel flat fading sub-channels in the frequency domain.

Each sub-channel is associated with a subcarrier. Let $h_{c^*,c,i}^n[k]$ denote the k-th sub-channel coefficient between the n-th user in the c-th cell and the i-th antenna of the BS of cell c^* in the uplink. The received signal $y_{c^*,i}$ by the i-th antenna element of the cell c^* at the k-th subcarrier can be expressed as

$$y_{c^*,i}[k] = \sum_{n=1}^{N} h_{c^*,c^*,i}^n[k] x_{c^*}^n[k] + \sum_{c=1,c\neq c^*}^{C} \sum_{n=1}^{N} h_{c^*,c,i}^n[k] x_c^n[k] + v_{c^*,i}[k], \tag{3.1}$$

for all $1 \leq i \leq M$ and $1 \leq c \leq C$, where $V_{c^*,i}[k]$ is the AWGN at the i-th antenna of the BS in cell c^* at the k-th subcarrier. Letting $\mathbf{y}_{c^*,i} = [y_{c*,i}[1] \cdots y_{c*,i}[K]]^T$, we can write (3.1) for all subcarriers at the i-th antenna of the BS in cell c^* in the compact form as

$$\mathbf{y}_{c^*,i} = \sum_{n=1}^{N} \mathbf{X}_{c^*}^n \mathbf{h}_{c^*,c^*,i}^n + \sum_{c=1,c\neq c^*}^{C} \sum_{n=1}^{N} \mathbf{X}_c^n \mathbf{h}_{c^*,c,i}^n + \mathbf{v}_{c^*,i}, \tag{3.2}$$

where $\mathbf{X}_{c^*}^n = \text{diag}\{\mathbf{x}_{c^*}^n\}$, $\mathbf{h}_{c^*,c,i}^n = [h_{c^*,c,i}^n[1] \cdots h_{c^*,c,i}^n[K]]^T$ and $\mathbf{v}_{c^*,i} = [v_{c^*,i}[1] \cdots v_{c^*,i}[K]]^T \sim \mathcal{CN}(0, \sigma_v^2)$. Let $\mathbf{g}_{c^*,c,i}^n = [g_{c^*,c,i}^n[1] \cdots g_{c^*,c,i}^n[\ell] \cdots g_{c^*,c,i}^n[L]]^T$ collect the samples of the sampled multipath channel impulse response (CIR) between the n-th user of the c-th cell and the i-th antenna of the BS in cell c^*, where L is the number of the channel taps and $g_{c^*,c,i}^n[\ell]$ corresponds to the ℓ-th channel tap. The K frequency domain channel coefficients, i.e., $\mathbf{h}_{c^*,c,i}^n$, can be calculated as the

K-point DFT of the CIR samples, i.e., $\mathbf{g}^{n}_{c^*,c,i} \in \mathbb{C}^{L\times 1}$, e.g., [60]. Hence,

$$\mathbf{h}^{n}_{c^*,c,i} = \mathbf{F}\mathbf{g}'^{n}_{c^*,c,i}, \tag{3.3}$$

where $\mathbf{F} \in \mathbb{C}^{K\times K}$ represents the discrete Fourier transform (DFT) matrix, whose element in row s and column r is given by $[\frac{1}{\sqrt{K}}e^{-j2\pi*(K-r)(K-s)/K}]$, $1 \leq r \leq K$ and $1 \leq s \leq K$ and $\mathbf{g}'^{n}_{c^*,c,i} \in \mathbb{C}^{K\times 1}$ is $\mathbf{g}^{n}_{c^*,c,i} \in \mathbb{C}^{L\times 1}$ augmented with $K-L$ zeros. Using (3.3) in (3.2), we get

$$\mathbf{y}_{c^*,i} = \sum_{n=1}^{N} \mathbf{X}^{n}_{c^*}\mathbf{F}\mathbf{g}'^{n}_{c^*,c^*,i} + \sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N} \mathbf{X}^{n}_{c}\mathbf{F}\mathbf{g}'^{n}_{c^*,c,i} + \mathbf{v}_{c^*,i}. \tag{3.4}$$

The channel coefficient is modelled as $g^{n}_{c^*,c,i}[\ell] = \sqrt{\phi_{c^*,c,i}}[\ell]\psi_{c^*,c,i}[\ell]$ for $1 \leq \ell \leq L$, where $\phi_{c^*,c,i}$ model the path-loss and shadowing (large-scale fading) that change slowly and can be learned over long period of time, while the term $\psi_{c^*,c,i}$ is assumed to be independent identical distribution (i.i.d) of unknown random variables with $\mathcal{CN}(0,1)$ (small-scale fading) [30]. Since the cell layout and shadowing are captured using the constant $\phi_{c^*,c,i}[\ell]$ values, for the purpose of this chapter, the specific details of the cell layout and shadowing model are irrelevant. In other words, any cell layout and any shadowing model can be incorporated with the above abstraction [17] and [30].

The received signal of (3.4) can be re-written as

$$\mathbf{y}_{c^*,i} = \sum_{n=1}^{N} \mathbf{X}^{n}_{c^*}\mathbf{F}\mathbf{g}'^{n}_{c^*,c^*,i} + \mathbf{z}_{c^*,i}, \tag{3.5}$$

where the term $\mathbf{z}_{c^*,i} = \sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N} \mathbf{X}^{n}_{c}\mathbf{F}\mathbf{g}'^{n}_{c^*,c,i} + \mathbf{v}_{c^*,i}$ in (3.5) represents the net sum of inter-cell interference plus the receiver noise, the variance interference σ_I^2 of the inter-cell interference term caused during pilot transmission can be

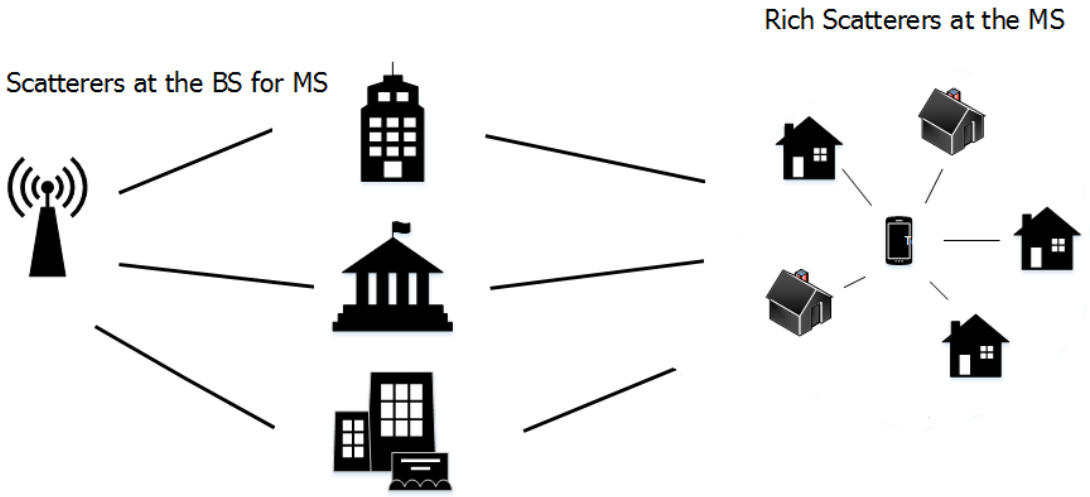

Figure 3-2: Illustration of the rich scatterers wireless channel and the resulting channel impulse response is sparse [61]

expressed as

$$\sigma_I^2 = E\{(\sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N}\mathbf{X}_c^n\mathbf{F}\mathbf{g}'^n_{c^*,c,i}) \times (\sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N}\mathbf{X}_c^n\mathbf{F}\mathbf{g}'^n_{c^*,c,i})^H\}. \tag{3.6}$$

We define the measurement matrix $\mathbf{A}_{c^*}^n = \mathbf{X}_{c^*}^n\mathbf{F}$, then (3.5) can be rewritten as

$$\mathbf{y}_{c^*,i} = \sum_{n=1}^{N}\mathbf{A}_{c^*}^n\mathbf{g}'^n_{c^*,c^*,i} + \mathbf{z}_{c^*,i}. \tag{3.7}$$

Based on the physical properties of outdoor electromagnetic propagation, the CIR in wireless communications usually contain a few significant channel taps as can be shown in Fig. 3-2, i.e. the CIR are sparse; hence, the number of non-zero taps of the channel is much smaller than the channel length, then the CS techniques can be applied for sparse channel estimation. This sparse property can be exploited to reduce the necessary channel parameters to be estimated. In this case, we can address the pilot contamination problem by using fewer pilots than the unknown channel coefficients [60]-[62].

3.2 BCS-Based Channel Estimation

In this section, BCS-based channel estimation is presented in the context of massive MIMO channel estimation. Following the general procedure of BCS in [63] and [64], the full posterior distribution over unknown parameters of interest for the problem at hand can be given as

$$P(\mathbf{g}'^{n}_{c^*,c^*,i}, \boldsymbol{\beta}, \sigma^2 | \mathbf{y}_{c^*,i}) = \frac{P(\mathbf{y}_{c^*,i} | \mathbf{g}'^{n}_{c^*,c^*,i}, \boldsymbol{\beta}, \sigma^2) P(\mathbf{g}'^{n}_{c^*,c^*,i}, \boldsymbol{\beta}, \sigma^2)}{P(\mathbf{y}_{c^*,i})}, \tag{3.8}$$

where $\boldsymbol{\beta}$ represents the hyperparameters[1] that control the sparsity of the channel while σ^2 is the net sum of the noise variance and interference variance.

However, the probability of the observation vector, $P(\mathbf{y}_{c^*,i})$, is defined by the following equation

$$P(\mathbf{y}_{c^*,i}) = \int\int\int P(\mathbf{y}_{c^*,i} | \mathbf{g}'^{n}_{c^*,c^*,i}, \sigma^2, \boldsymbol{\beta}) P(\mathbf{g}'^{n}_{c^*,c^*,i}, \boldsymbol{\beta}, \sigma^2) d\mathbf{g}' \; d\boldsymbol{\beta} \; d\sigma^2, \tag{3.9}$$

cannot be computed analytically. To go around that, an approximation strategy must be applied. The posterior distribution can be decomposed based on the product rule of productivity as

$$P(\mathbf{g}'^{n}_{c^*,c^*,i}, \boldsymbol{\beta}, \sigma^2 | \mathbf{y}_{c^*,i}) \equiv P(\mathbf{g}'^{n}_{c^*,c^*,i} | \mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2) P(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}_{c^*,i}). \tag{3.10}$$

The first term of (3.10), $P(\mathbf{g}'^{n}_{c^*,c^*,i} | \mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2)$, the posterior distribution over the channel coefficient can be expressed based on Bayes' rule as

$$P(\mathbf{g}'^{n}_{c^*,c^*,i} | \mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2) = \frac{P(\mathbf{y}_{c^*,i} | \mathbf{g}'^{n}_{c^*,c^*,i}, \sigma^2) P(\mathbf{g}'^{n}_{c^*,c^*,i} | \boldsymbol{\beta})}{P(\mathbf{y}_{c^*,i} | \boldsymbol{\beta}, \sigma^2)}. \tag{3.11}$$

The posterior distribution given above is Gaussian distribution with mean

[1] a parameter of a prior distribution

$\boldsymbol{\mu}^n_{c^*,c^*,i}$ and the variance $\boldsymbol{\Sigma}^n_{c^*,c^*,i}$ are given by

$$\boldsymbol{\mu}^n_{c^*,c^*,i} = \sigma^{-2}\boldsymbol{\Sigma}^n_{c^*,c^*,i}\mathbf{A}^n_{c^*}\mathbf{y}_{c^*,i}, \tag{3.12}$$

$$\boldsymbol{\Sigma}^n_{c^*,c^*,i} = (\boldsymbol{\zeta} + \sigma^{-2}(\mathbf{A}^n_{c^*})^T\mathbf{A}^n_{c^*})^{-1}, \tag{3.13}$$

where $\boldsymbol{\zeta} = diag\{\beta_1, \beta_2, ..., \beta_K\}$.

The estimated channel based on Bayesian estimation approaches to minimize the mean square error (MSE) is the expectation of $P(\mathbf{g}'^n_{c^*,c^*,i}|\mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2)$, so the estimated channel can be expressed as

$$\hat{\mathbf{g}}'^n_{c^*,c^*,i} = E(P(\mathbf{g}'^n_{c^*,c^*,i}|\mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2)) = \boldsymbol{\mu}^n_{c^*,c^*,i}. \tag{3.14}$$

Now, to obtain the estimated channel $\hat{\mathbf{g}}'^n_{c^*,c^*,i}$, we need to find the heypar-marpater σ^2 and $\boldsymbol{\beta}$ that can be obtained from the second term on the right-hand side of (3.10) by applying a type$-II$ maximum likelihood procedure.

Based on Bayes' theorem, the posterior distribution $P(\boldsymbol{\beta}, \sigma^2|\mathbf{y}_{c^*,i})$ is proportional $P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta}, \sigma^2)$ [42], Then, the type$-II$ maximum likelihood is applied to the log marginal likelihood as follows

$$P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta}, \sigma^2) = \int_{-\infty}^{\infty} P(\mathbf{y}_{c^*,i}|\mathbf{g}'^n_{c^*,c^*,i}, \sigma^2)P(\mathbf{g}'^n_{c^*,c^*,i}|\boldsymbol{\beta})d\mathbf{g}'. \tag{3.15}$$

The term $P(\mathbf{g}'^n_{c^*,c^*,i}|\boldsymbol{\beta})$ follows zero-mean Gaussian distribution and can be expressed as

$$P(\mathbf{g}'^n_{c^*,c^*,i}|\boldsymbol{\beta}) = (2\pi)^{\frac{-K}{2}} \prod_{i=1}^{K} \beta_k^{\frac{1}{2}} exp[\frac{-1}{2}\mathbf{g}'^n_{c^*,c^*,i}\boldsymbol{\zeta}(\mathbf{g}'^n_{c^*,c^*,i})^H], \tag{3.16}$$

while the Gaussian likelihood function of $\mathbf{y}_{c^*,i}$ according to the probability theory, can be written as

$$P(\mathbf{y}_{c^*,i}|\mathbf{g}'^n_{c^*,c^*,i},\sigma^2) = (\frac{2\pi}{\sigma^2})^{\frac{-K}{2}} exp(\frac{-\sigma^2}{2}||\mathbf{y}_{c^*,i} - \mathbf{A}^n_{c^*}\mathbf{g}'^n_{c^*,c^*,i}||_2^2). \tag{3.17}$$

By substituting (3.16) and (3.17) into (3.15), marginal likelihood $P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta},\sigma^2)$ can be expressed as

$$P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta},\sigma^2) = \log\{(\frac{\beta_k}{2\pi})^{\frac{K}{2}}(\frac{1}{2\pi})^{\frac{K}{2}}\prod_{k=1}^{K}\beta_k^{\frac{1}{2}}\int_{-\infty}^{\infty}\exp(-(\frac{\beta_k}{2}||\mathbf{y}_{c^*,i} - \mathbf{A}^n_{c^*}\mathbf{g}'^n_{c^*,c^*,i}||_2^2) + \frac{1}{2}(\mathbf{g}'^n_{c^*,c^*,i})^H\boldsymbol{\zeta}\mathbf{g}'^n_{c^*,c^*,i}))\}, \tag{3.18}$$

$\boldsymbol{\beta}$ can be obtained by differentiating the log marginal likelihood $P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta},\sigma^2)$ with regard to σ^2, and equating it to zero and it can be given as

$$(\beta_k)^{ii} = \frac{I - \beta_k(\Sigma^n_{c^*,c^*,i})_k}{(\mu^n_{c^*,c^*,i})_k^2}. \tag{3.19}$$

While σ^2 is obtained by differentiate (3.18) with regard to $\boldsymbol{\beta}$ and set these derivations to zero and can be expressed as

$$(\sigma^2)^{ii} = \frac{||\mathbf{y}_{c^*,i} - \mathbf{A}^n_{c^*}\mathbf{g}'^n_{c^*,c^*,i}||_2^2}{(M - I + \sum_{k=1}^{K}\beta_k)}. \tag{3.20}$$

The β_k and σ_k^2 which maximize the log marginal likelihood $P(\mathbf{y}_{c^*,i}|\boldsymbol{\beta},\sigma^2)$ are then found iteratively by setting $\boldsymbol{\beta}$ and σ^2 to initial values and then finding values for $\boldsymbol{\mu}^n_{c^*,c^*,i}$ and $\boldsymbol{\Sigma}^n_{c^*,c^*,i}$ from (3.12) and (3.13). These values are then repeatedly used to calculate a new estimate for β_k and σ^2 and until a convergence criteria is met.

Further details of the BCS algorithm can be found in [63], [64]. The procedure

for implementation of the proposed technique is summarized in algorithm 1.

In contrast to the conventional BCS-based estimator, it can also improve the performance of the BCS estimator based on the principle of thresholding, which can be applied to keep the most significant taps. The proposed algorithm applies a threshold approach by retaining the channel taps that have energy above a threshold value of ϱ and set the other taps to zero. The value of ϱ is the energy of the channel impulse response.

Algorithm 1 BCS-Based Channel Estimation Algorithm

INPUTS:

1: Pilot signal $\mathbf{x}_{c^*}^n$.

2: Observation matrix $\mathbf{A}_{c^*}^n = \mathbf{x}_{c^*}^n \mathbf{F}$.

Initial Configuration:

3: Select a suitable value for convergence δ.

4: $ii \leftarrow 1$

5: Select a start value for σ^2 and $\boldsymbol{\beta}$.

6: Compute $\boldsymbol{\Sigma}_{c^*,c^*,i}^n = (\boldsymbol{\zeta} + \boldsymbol{\beta}(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)^{-1}$.

7: Compute $\boldsymbol{\mu}_{c^*,c^*,i}^n = \boldsymbol{\beta}\boldsymbol{\Sigma}_{c^*,c^*,i}^n \mathbf{A}_{c^*}^n \mathbf{y}_{c^*,i}$.

repeat

8:Update $(\beta_k)^{ii} = \frac{I-\beta_k(\Sigma_{c^*,c^*,i}^n)_k}{(\mu_{c^*,c^*,i}^n)_k^2}$ and $(\sigma^2)^{ii} = \frac{||\mathbf{y}_{c^*,i} - \mathbf{A}_{c^*}^n \mathbf{g'}_{c^*,c,i}^n||_2^2}{(M-I+\sum_{k=1}^K \beta_k)}$.

9: **Until** $\delta > \Sigma_{ii=1}(\sigma^2)^{ii+1} - (\sigma^2)^{ii}$.

10: Compute $\hat{\mathbf{g}}_{c^*,c^*,i}^{\prime n} = \boldsymbol{\mu}_{c^*,c^*,i}^n = E(P(\mathbf{g}_{c^*,c^*,i}^{\prime n} | \mathbf{y}_{c^*,i}, \boldsymbol{\beta}, \sigma^2))$.

OUTPUTS: Return the Estimated Channel $\hat{\mathbf{g}}_{c^*,c^*,i}^{\prime n}$.

3.3 Multi-Task BCS Based Channel Estimation

With a high probability of user movements, the massive MIMO system channel may vary. Consequently, the channels at different time instants/locations are different but share the same common statistical property. As a result, to estimate the current channel, we can exploit the previous compressive vectors in addition to the current compressive vector [66].

Given the system model in II, the received signals of (3.7) can have the following formulation

$$\mathbf{y}_{c^*,i,j} = \sum_{n=1}^{N} \mathbf{A}^n_{c^*,j} \mathbf{g}'^n_{c^*,c^*,i,j} + \mathbf{z}_{c^*,i,j}, \tag{3.21}$$

for $j = 1, 2, ...J$ where J is the number of the task, $\mathbf{A}^n_{c^*,j}, \mathbf{g}'^n_{c^*,c^*,i,j}$ and $\mathbf{z}_{c^*,i,j}$ represents the jth measurement matrices, channel vector and the noise vector, respectively [66].

The main target is to estimate the channel $\mathbf{g}'^n_{c^*,c^*,i,j}$ which can be computed based on Bayesian channel estimation philosophy as the mean of the channel posterior distribution that can be represented as

$$\hat{\mathbf{g}}'^n_{c^*,c^*,i,j} = E(P(\mathbf{g}'^n_{c^*,c^*,i,j} | \mathbf{y}_{c^*,i,j}, \boldsymbol{\Xi}_j, \xi_0)), \tag{3.22}$$

where ξ_0 represents the inverse of the net sum of the noise variance and interference variance, while $\boldsymbol{\Xi}_j$ represent the hyperparameters that control the sparsity of the channel. Based on Bayes' rule the posterior distribution can be given as

$$P(\mathbf{g}'^n_{c^*,c^*,i,j} | \mathbf{y}_{c^*,i,j}, \boldsymbol{\Xi}_j, \xi_0)$$

$$= \frac{P(\mathbf{y}_{c^*,i,j} | \mathbf{g}'^n_{c^*,c^*,i,j}, \xi_0) P(\mathbf{g}'^n_{c^*,c^*,i,j} | \boldsymbol{\Xi}_j)}{\int P(\mathbf{y}_{c^*,i,j} | \mathbf{g}'^n_{c^*,c^*,i,j}, \xi_0) P(\mathbf{g}'^n_{c^*,c^*,i,j}, \boldsymbol{\Xi}_j) d\mathbf{g}'}$$

$$\sim N(\boldsymbol{\mu}^n_{c^*,i,j}, \boldsymbol{\Sigma}^n_{c^*,i,j}), \tag{3.23}$$

the mean and covariance can be given by

$$\boldsymbol{\mu}^n_{c^*,i,j} = \xi_0 \boldsymbol{\Sigma}^n_{c^*,i,j} \mathbf{A}^n_{c^*,j} \mathbf{y}_{c^*,i,j}, \tag{3.24}$$

$$\boldsymbol{\Sigma}^n_{c^*,i,j} = (\boldsymbol{\psi} + \boldsymbol{\Xi}_j (\mathbf{A}^n_{c^*,j})^T \mathbf{A}^n_{c^*,j})^{-1}, \tag{3.25}$$

where $\boldsymbol{\psi} = diag(\psi_0, \psi_1, \psi_2, ..., \psi_K)$.

The likelihood function for the parameter $\mathbf{g}'^{n}_{c^*,c^*,i,j}$ and ξ_0 based on the received signal $\mathbf{y}_{c^*,i,j}$ and can be expressed as

$$P(\mathbf{y}_{c^*,i,j}|\mathbf{g}'^{n}_{c^*,c^*,i,j},\xi_0) = (\frac{2\pi}{\xi_0})^{\frac{-N}{2}} \exp(\frac{-\xi_0}{2}||\mathbf{y}_{c^*,i,j} - \mathbf{A}^{n}_{c^*,j}\mathbf{g}'^{n}_{c^*,c^*,i,j}||_2^2). \tag{3.26}$$

The channel coefficients $\mathbf{g}'^{n}_{c^*,c^*,i,j}$ are assumed to be drawn from a product of zero-mean Gaussian distributions that are shared by all tasks as follow

$$\begin{aligned} P(\mathbf{g}'^{n}_{c^*,c^*,i,j}|\boldsymbol{\Xi}_j) =& \prod_{i=1}^{N} (\mathbf{g}'^{n}_{c^*,c^*,i,j}|0,\boldsymbol{\Xi}_j^{-1}) = (2\pi)^{\frac{-N}{2}} \prod_{i=1}^{N} \boldsymbol{\Xi}_j^{\frac{1}{2}} \\ & \times \exp[\frac{-1}{2}(\mathbf{g}'^{n}_{c^*,c^*,i,j})^H \boldsymbol{\Xi}_j \mathbf{g}'^{n}_{c^*,c^*,i,j}]. \end{aligned} \tag{3.27}$$

To obtain the estimated channel, we need to estimate $\boldsymbol{\Xi}_j$ and ξ_0 by applying the same procedure in section 3.2 to the posterior distribution $P(\mathbf{y}_{c^*,i,j}|,\boldsymbol{\Xi}_j,\xi_0)$ that can be inferred as [66]

$$P(\mathbf{y}_{c^*,i,j}|\boldsymbol{\Xi}_j,\xi_0) \equiv P(\mathbf{y}_{c^*,i,j}|\mathbf{g}'^{n}_{c^*,c^*,i,j},\xi_0) P(\mathbf{g}'^{n}_{c^*,c^*,i,j}|\boldsymbol{\Xi}_j). \tag{3.28}$$

Now, by maximizing the log marginal likelihood and then differentiating with respect to $\boldsymbol{\Xi}_j$ and ξ_0 and setting to zero yields

$$(\boldsymbol{\Xi}_j)^{new} = \frac{J - \boldsymbol{\Xi}_j \sum_{j=1}^{J} \boldsymbol{\Sigma}^{n}_{c^*,c^*,i,j}}{\sum_{j=1}^{J} (\boldsymbol{\mu}^{n}_{c^*,c^*,i,j})^2}, \tag{3.29}$$

$$(\xi_0)^{new} = \frac{\sum_{j=1}^{J} (K - J + \sum_{i=1}^{J} \boldsymbol{\Sigma}^{n}_{c^*,c^*,i,j} \boldsymbol{\Xi}_j)}{\sum_{j=1}^{J} ||\mathbf{y}_{c^*,i,j} - \mathbf{A}^{n}_{c^*,j} \mathbf{g}'^{n}_{c^*,c^*,i,j}||_2^2}. \tag{3.30}$$

Further information on MT-BCS can be found in [67].

3.4 CRB For BCS-Based Estimator

In this section, we analyse the CRB for the proposed BCS and MT-BCS based channel estimation techniques to provide a benchmark for the minimum estimation error that can be achieved by the proposed algorithm. The CRB on the covariance of any estimator $\hat{\boldsymbol{\theta}}$ can be given as

$$E\{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^T\} \geq J^{-1}(\boldsymbol{\theta}), \tag{3.31}$$

where $J(\boldsymbol{\theta})$ is the Fisher information matrix (FIM) corresponding to the observation f, and can be given as

$$J(\boldsymbol{\theta}) = E(\frac{\partial}{\partial \boldsymbol{\theta}} \log l(\boldsymbol{\theta}, f))(\frac{\partial}{\partial \boldsymbol{\theta}} \log l(\boldsymbol{\theta}, f))^T, \tag{3.32}$$

where $l(\boldsymbol{\theta}, f)$ is the likelihood function corresponding to the observation f, parameterized by $\boldsymbol{\theta}$ [67].

Based on (3.32), we can write the FIM as

$$J(\mathbf{y}_{c^*,i,j}) \geq -E(\frac{\partial^2 log(P_{\mathbf{y}_{c^*,i,j}|\boldsymbol{\Xi}_j,\xi_0}(P(\mathbf{y}_{c^*,i,j}|\boldsymbol{\Xi}_j,\xi_0)))}{\partial^2 \mathbf{g}'})^{-1} \tag{3.33}$$

Reference the Bayes' rule in (3.23), the FIM can be decomposed into two terms

$$\begin{aligned} &- E(\frac{\partial^2 log(P_{\mathbf{y}_{c^*,i,j}|\boldsymbol{\Xi}_j,\xi_0}(P(\mathbf{y}_{c^*,i,j}|\boldsymbol{\Xi}_j,\xi_0)))}{\partial^2 \mathbf{g}'}) = - \\ &E(\frac{\partial^2 log(P_{\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0}(P(\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0))}{\partial^2 \mathbf{g}'})- \\ &E(\frac{\partial^2 log(P_{\mathbf{g}'^n_{c^*,c^*,i,j}|\boldsymbol{\Xi}_j}(P(\mathbf{g}'^n_{c^*,c^*,i,j}|\boldsymbol{\Xi}_j)))}{\partial^2 \mathbf{g}'}), \end{aligned} \tag{3.34}$$

which can be expressed in matrix form as

$$\mathbf{J} = \mathbf{J}_D + \mathbf{J}_P, \tag{3.35}$$

where $\mathbf{J}$, $\mathbf{J}_D$ and $\mathbf{J}_P$ represent the Bayesian FIM, data information matrix and prior information matrix, respectively.

Using (3.26), the data information matrix $\mathbf{J}_D$ can be given as

$$\mathbf{J}_D = -\frac{\partial^2 log(P_{\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0}(P(\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0))}{\partial^2 \mathbf{g}'} =$$

$$\frac{\partial}{\partial \mathbf{g}'}[-log(2\pi)^{\frac{1}{2}}\xi_0^{-1} - \frac{\xi_0}{2}||\mathbf{y}_{c^*,i,j} - \mathbf{A}^n_{c^*,j}\mathbf{g}'^n_{c^*,c^*,i,j}||_2^2], \tag{3.36}$$

$$= \frac{1}{\xi_0}(\mathbf{y}_{c^*,i,j} - \mathbf{A}^n_{c^*,j}\mathbf{g}'^n_{c^*,c^*,i,j}), \tag{3.37}$$

by applying the second derivative to (3.37), we get

$$\frac{\partial^2 log(P_{\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0}(P(\mathbf{y}_{c^*,i,j}|\mathbf{g}'^n_{c^*,c^*,i,j},\xi_0))}{\partial^2 \mathbf{g}'} = \frac{\mathbf{A}^n_{c^*,j}(\mathbf{A}^n_{c^*,j})^T}{\xi_0}. \tag{3.38}$$

By applying the same procedure in (3.36-3.38) to the second term of (3.34) gives

$$\mathbf{J}_D = E(\frac{\partial^2 log(P_{\mathbf{g}'^n_{c^*,c^*,i,j}|\mathbf{\Xi}_j}(P(\mathbf{g}'^n_{c^*,c^*,i,j}|\mathbf{\Xi}_j)))}{\partial^2 \mathbf{g}'}) = (\mathbf{\Xi}_j)^{-1}. \tag{3.39}$$

Thus, the closed form expression of the BCRB for the proposed MT-BCS can

be given as

$$J(\mathbf{g}'^{n}_{c^*,c^*,i,j}) \geq (\frac{1}{\mathbf{\Xi}_j} + \frac{\mathbf{A}^n_{c^*,j}(\mathbf{A}^n_{c^*,j})^T}{\xi_0})^{-1}. \tag{3.40}$$

3.5 Simulation Results

To verify the accuracy of our analytical results, we consider a multi cell system with 7 cells, each cell has a base station equipped with $M = 100$ antennae and serves $N = 10$ users. The number of the channel taps L is 128, the number of subcarrier K is 4096 and the convergence δ is 10^{-6}. The simulation results are obtained by averaging over 1000 realizations.

To compare the accuracy of the channel estimation techniques, the normalized MSE is used for performance evaluation and is computed as

$$MSE = \frac{||\hat{\mathbf{g}}'^{n}_{c^*,c^*,i,j} - \mathbf{g}'^{n}_{c^*,c^*,i,j}||_2^2}{||\mathbf{g}'^{n}_{c^*,c^*,i,j}||_2^2}. \tag{3.41}$$

Fig. 3-3 shows the MSE performance comparison among a BCS-based channel estimation of three scenarios under small pilot contamination ($\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c,i} = 0.1$), strong pilot contamination ($\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c,i} = 0.5$), very strong pilot contamination ($\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c,i} = 0.9$), regularized least square (RLS)-based estimator with no pilot contamination as a benchmark and the BCRB for BCS as a reference line. The results have shown significant improvement in estimation accuracy and addressing the pilot contamination problem for SNR values of -40 dB to 40 dB for the proposed technique compared with R-LS. This is a result of exploiting the prior statistical of channel sparsity. Furthermore, the results still show enhanced estimation performance for high SNR.

Fig. 3-4 shows the MSE performance versus SNR with a different value of

setting to the number of subcarrier $K = \{640, 1280 \text{ and } 1920\}$, so the compression ratio (CR) (i.e. L/K) is to be CR= $\{0.2, 0.1\text{and } 0.06\}$, while the experiment is run under small pilot contamination ($\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c,i} = 0.1$). The results prove that the estimation accuracy is better performed by decreasing the values of the number of subcarriers, accordingly with increasing CR.

Fig. 3-5 demonstrates the MSE of the BSC-based channel estimation versus SNR for three scenarios of different settings to the number of antennae at the base station $M = \{100, 200 \text{ and } 300\}$, the system under strong pilot contamination ($\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c,i} = 0.7$). The results show that the estimation accuracy of the proposed algorithm is enhanced by increasing the number of antennae. Thus according to the law of large numbers, more coordinated BS antennae could provide more accurate support estimation.

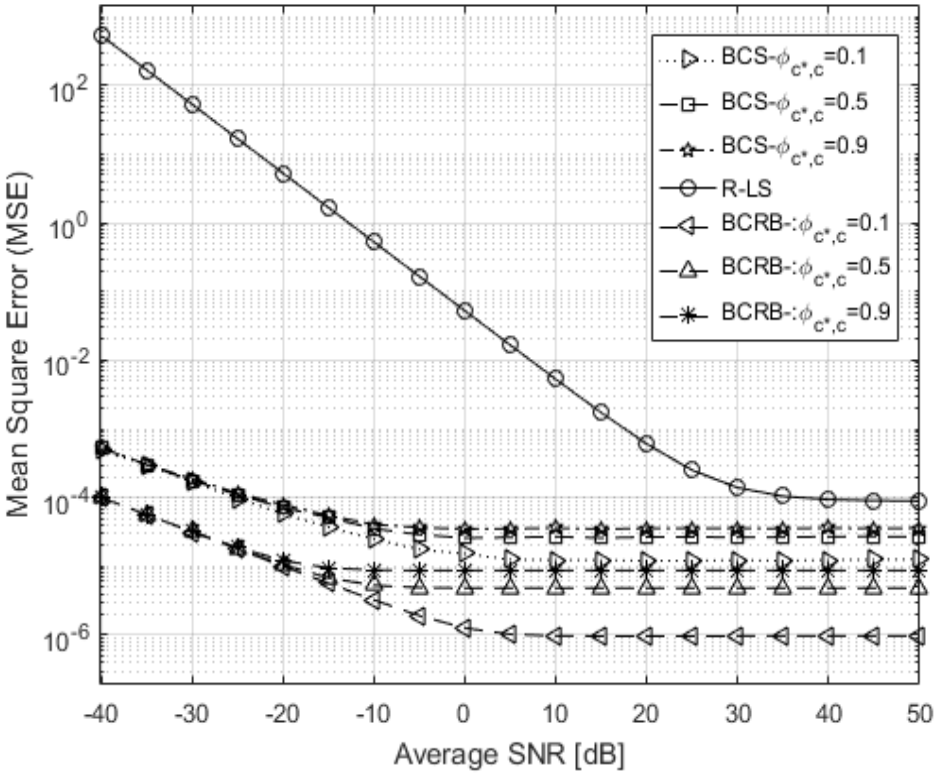

Figure 3-3: MSE performance comparison between BSC, BCRB for $\phi_{c*,c,i} = \{0.1, 0.5, 0.9\}$ and R-LS versus SNR.

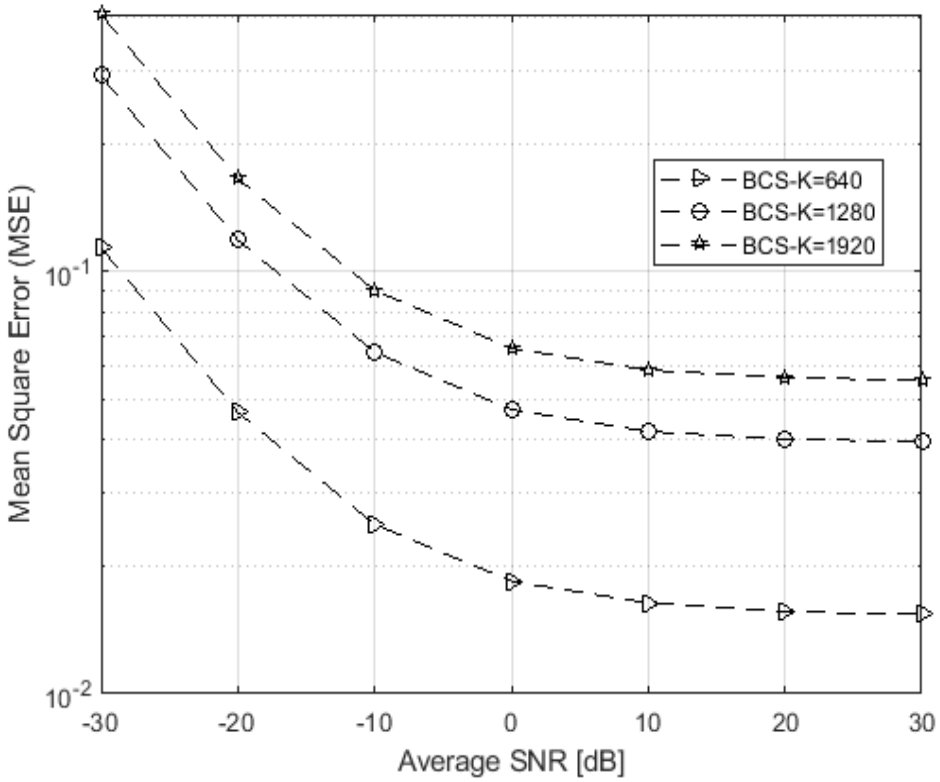

Figure 3-4: MSE of BSC for $K = \{640, 1280 \text{ and } 1920\}$ and $CR = \{0.2, 0.1 \text{ and } 0.06\}$, respectively.

Fig. 3-6 shows the MSE performance versus SNR for BCS with different values for the number of pilots: 1000, 500, 100, 50 and 25, where the number of subcarrier K is 4096. The number of the CIR path is 128 while the experiments run under strong pilot contamination. For cases of the number of the pilots is greater than the number of channel taps (i.e., 1000 and 500), the BCS provides inefficient estimation accuracy, while for the other cases of the number of the pilot of (100, 50 and 25), which is less than 128, the estimation accuracy is enhanced significantly. In addition, there is no significant improvement for the cases of the number of the pilots 100, 50 and 25. In these cases, we can address pilot contamination by employing small values for the number of the pilot i.e. 25.

Fig. 3-7 compares the MSE performance versus SNR among BCS, threshold-BSC, MT-BCS, LS, orthogonal message passing (OMP) and the bilinear approx-

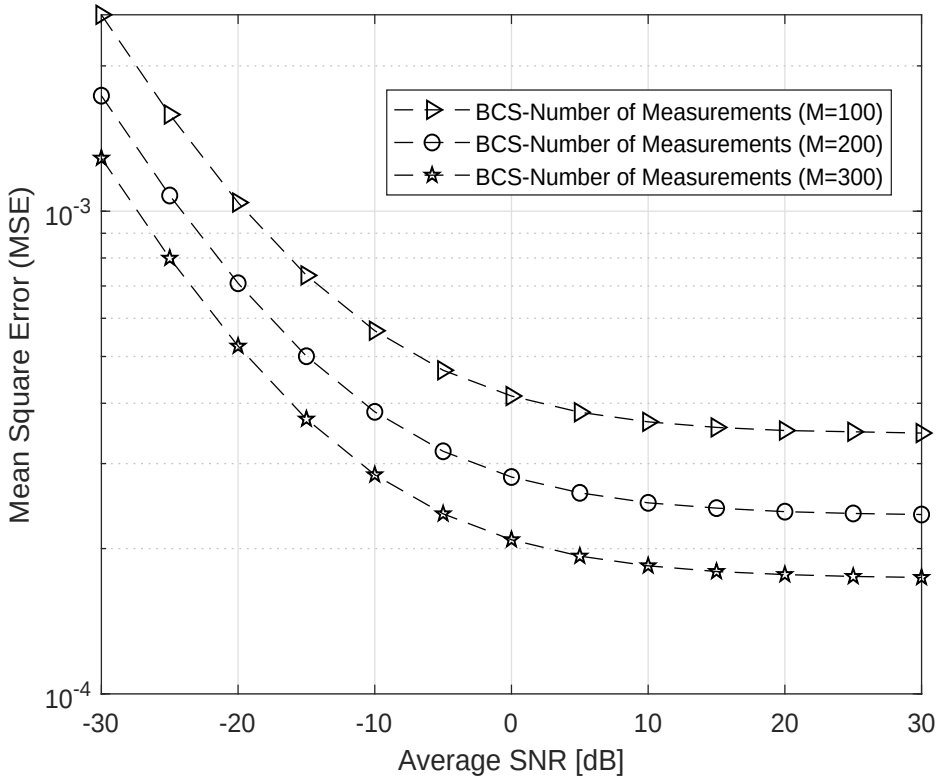

Figure 3-5: MSE of BSC for $M = \{100, 200 \text{ and } 300\}$ versus SNR.

Table 3.1: Complexity Analysis

Estimators	Computation Complexity
R-LS	$O(L^2)$
BCS	$O(KL^2)$
MTBCS	$O(KL^3)$
OMP	$O(L\log(K))$
BiAMP	$O(LK + K)$

imate message passing (Bi-AMP)[43]. The number of subcarrier K is 1024 and the number of the CIR path is 100. Results show the proposed MT-BCS enjoys significant performance improvement over all the other estimators as a result of exploiting the statistical prior information on a large scale. However, this advantage is at the expense of a relatively high complexity of BCS and MT-BCS over other estimators as depicted in Table 1, which compares the computational

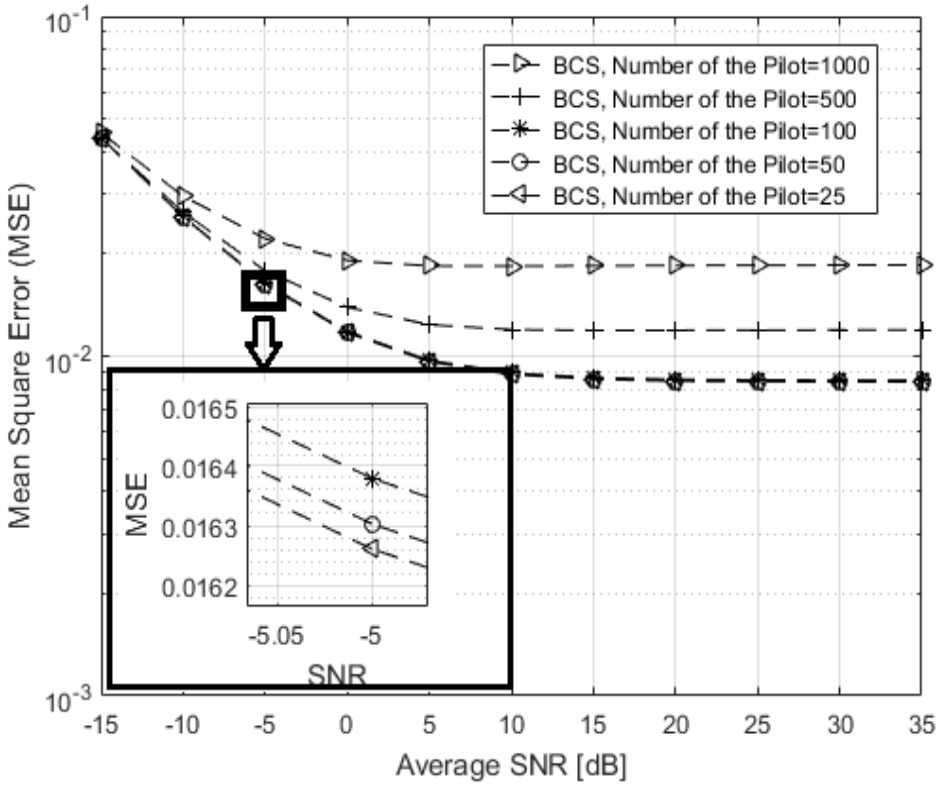

Figure 3-6: MSE performance comparison of BSC based estimator for different values of the number of the pilot $100, 50$ and 10 versus SNR.

complexity Bi-AMP [68], BCS [64], OMP [69], LS [70] and the MT-BCS [66]. Also, the results showed that the thresholding approach enhances the estimation accuracy of the conventional BCS, as the CIR contains so many taps with no significant energy. By setting the threshold and neglecting these taps, a huge part of the noise and interference from pilot contamination will be eliminated.

3.6 Conclusion

To address the pilot contamination problem in massive MIMO systems, we proposed a BCS-based channel estimation algorithm for the multi-cell multi-user massive MIMO. The simulation results have revealed that the BCS-based channel estimation algorithm has tremendous improvement over conventional-based

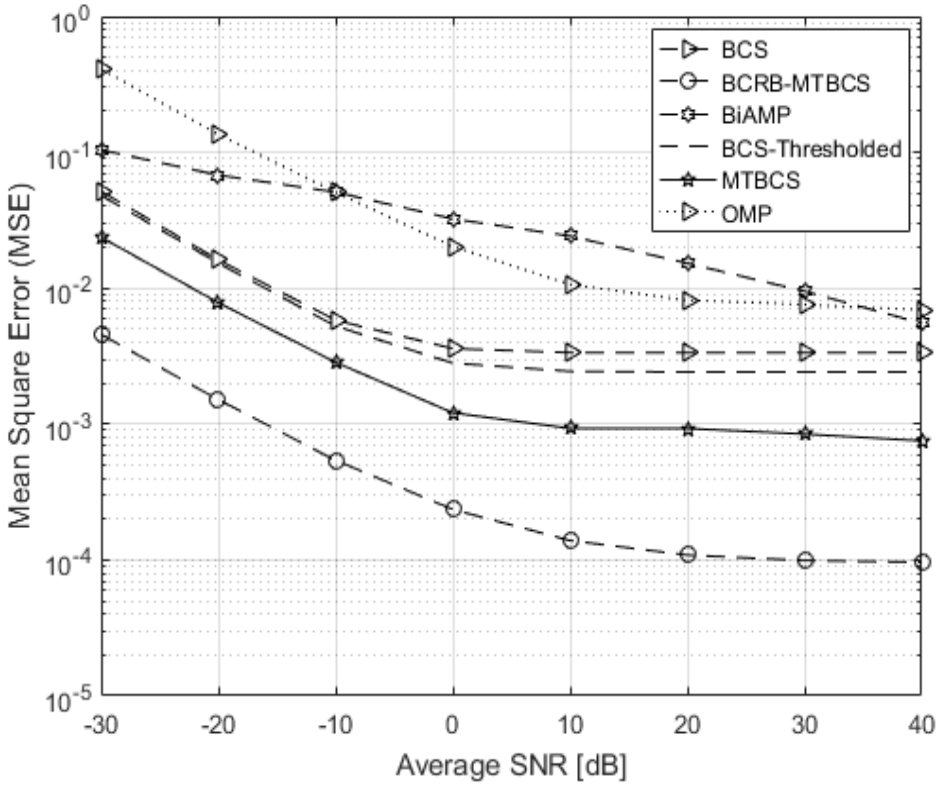

Figure 3-7: MSE performance comparison between BCS, Thresholded BCS, LS, MT-BCS, OMP and BiAMP based estimators versus SNR.

channel estimation algorithms and can address the pilot contamination problem. Furthermore, the proposed technique can be enhanced by thresholding the CIR to a certain value and also by exploiting the common sparsity feature inherent in the system channel. In addition, the number of antennae and the compression ratio should be selected wisely to achieve optimum estimation accuracy.

Chapter 4

Sparse Bayesian Learning Channel Estimation with Optimal Pilot Design

In the previous chapter, we proposed a Bayesian compressed sensing (BCS) approach to overcome the shortages of compressed sensing (CS) techniques. However, these works assume dependency among antennae elements. The antennae in MIMO systems are not well separated in realistic environments, so the MIMO channel is correlated [71]. Thus, in this chapter, we propose an improved channel estimation technique based on a sparse Bayesian learning (SBL) scheme that consider the impact of antennae correlation, namely, a pattern-coupled SBL [72]-[73], whereby a priori probabilistic information regarding channel sparsity is controlled by its own hyperparameter and its neighbouring hyperparameters. This dependency feature can be exploited to provide more reliable channel recovery. The simulation results show that the channel coefficients can be estimated more efficiently in contrast to the conventional channel estimators in terms of channel estimation with pilot contamination based on a mean square error (MSE) analytical expression. Furthermore, based on that MSE expression, a pilot design

criteria is formulated to design the optimal pilot to improve the estimation accuracy of the proposed algorithm. Results show that we can reduce the MSE of the SBL estimator by employing the optimal pilot sequence.

The remainder of this chapter is organized as follows. Section 4.1 describes the multi-cell massive MIMO system model. Section 4.2 analyses the SBL-based Channel Estimator. Optimal Pilot design analysis is presented in section 4.3. Section 4.4 provides the analysis of the achievable uplink rate. Section 4.5 presents the CRB for the proposed technique. Section 4.6 presents the simulation results and we summarize the conclusions in Section 4.7.

4.1 System Model

The system model considered in this chapter is the same model in the previous chapter with the exception of assuming that the channel impulse response (CIR) of different transmit-receive antenna pairs share a common sparse pattern for $\frac{d}{c} \leq \frac{1}{10B}$, where d is the distance between two antennae and c is the speed of light and B is the bandwidth of the signal bandwidth [74] and [75], we refer the interested reader to [76] for more clarifications about these assumption. Therefore, CIR of different transmit-receive antenna pairs share very similar scatters, as shown in Fig. 4-1. This sparsity communality of the channel is generalized to a block-sparse channel, in which a group of channel coefficient is sharing the same sparsity pattern. To take the practical advantages of block sparsity, the block structure CS framework is exploited by applying the pattern coupled-SBL approach to achieve better estimation accuracy.

4.2 SBL-Based Channel Estimator

In this section, the pattern-coupled sparse Bayesian learning method is presented in the context of massive MIMO channel estimation. Based on Bayesian channel

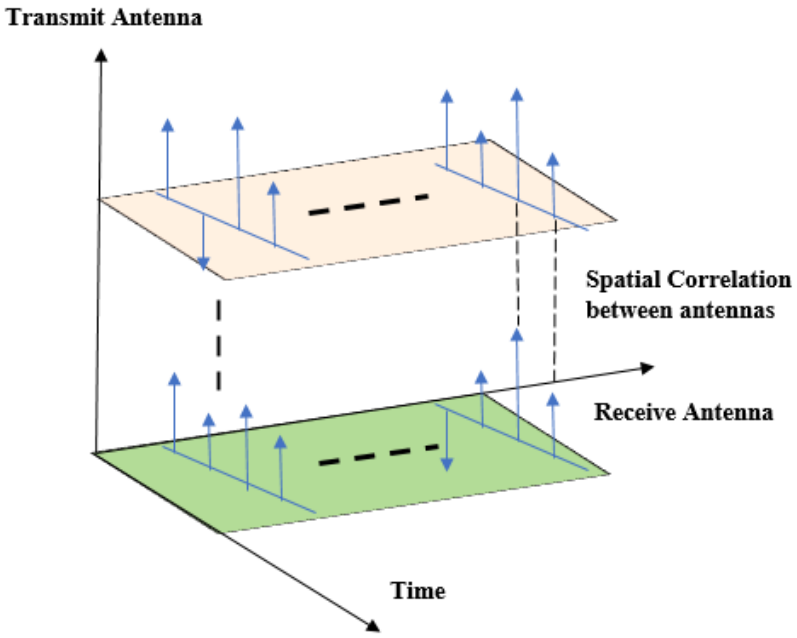

Figure 4-1: Spatial correlations of sparse MIMO channels.

estimation philosophy, the unknown parameters of interest can be estimated by applying the expectation of the posterior probability. As such, to obtain the estimated channel, we need to infer the posterior probability of the unknown parameters.

4.2.1 Bayesian Inference Model

Following the pattern-coupled sparse Bayesian learning model and based on Bayes' rule [72], [73], the full posterior distribution of $\mathbf{g}'^{n}_{c^*,c^*,i}$ over unknown parameters of interest for the problem at hand is proportional to the prior probability and the likelihood of the unknown parameters, that can be computed as

$$P(\mathbf{g}'^{n}_{c^*,c^*,i}|\boldsymbol{\alpha},\gamma,\mathbf{y}_{c^*,i}) = P(\mathbf{g}'^{n}_{c^*,c^*,i}|\boldsymbol{\alpha})P(\mathbf{y}_{c^*,i}|\mathbf{g}'^{n}_{c^*,c^*,i}), \tag{4.1}$$

where γ represents the inverse of the net sum of the noise and interference covariance matrices and $\boldsymbol{\alpha}$ are non-negative hyperparameters controlling the sparsity of the channel $\mathbf{g}'^{n}_{c^*,c^*,i}$. According to probability theory, the term $P(\mathbf{y}_{c^*,i}|\mathbf{g}'^{n}_{c^*,c^*,i})$ can be written as

$$P(\boldsymbol{y}_{c^*,i}|\mathbf{g}'^{n}_{c^*,c^*,i}) = (\frac{1}{\sqrt{2\pi}\gamma^{-1}}) \exp(-\frac{||\boldsymbol{y}_{c^*,i} - \mathbf{g}'^{n}_{c^*,c^*,i}\boldsymbol{A}^{n}_{c^*}||_2^2}{2\gamma^{-1}}), \tag{4.2}$$

The statistical properties of the sparse multipath structure of the channel is following Gaussian distribution based on the central line theorem [77]. So, the Gaussian prior for each channel coefficient $P(\mathbf{g}'^{n}_{c^*,c^*,i}|\boldsymbol{\alpha})$ in the pattern-coupled model is given by [72]

$$P(\mathbf{g}'^{n}_{c^*,c^*,i}|\boldsymbol{\alpha}) = \prod_{i=1}^{M} P(H^{n}_{c^*,c^*,i}[v,k]|\alpha_i, \alpha_{i+1}, \alpha_{i-1}) \tag{4.3}$$

$$\begin{aligned} =&(2\pi)^{\frac{-M}{2}} \prod_{i=1}^{M} (det(\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}))^{\frac{-M}{2}} \\ &\exp[\frac{-1}{2}(H^{n}_{c^*,c^*,i}[k])^T(\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1})^{-1} H^{n}_{c^*,c^*,i}[k]], i = 1, ..., M \end{aligned} \tag{4.4}$$

where $0 \leq \beta \leq 1$ is a parameter indicating the pattern relevance between the channel coefficient $H^{n}_{c^*,c^*,i}[k]$ and its neighboring coefficients $\{H^{n}_{c^*,c^*,i+1}[k], H^{n}_{c^*,c^*,i-1}[k]\}$. For $\beta = 0$, the Gaussian prior distribution in (4.4) is reduced to the prior for the conventional sparse Bayesian learning (which represents the uncorrelated channel scenario). The prior conjugate of Gaussian distribution times Gaussian distribution is a Gaussian distribution so, the posterior $P(\mathbf{g}'^{n}_{c^*,c,i}|\boldsymbol{\alpha}, \gamma, \boldsymbol{y}_{c^*,i}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ follows a Gaussian distribution with its mean and covariance given respectively by

$$\boldsymbol{\mu} = \gamma\boldsymbol{\Sigma}\mathbf{A}^{\mathbf{n}\,T}_{\mathbf{c}^*}\mathbf{y}_{c^*,i}, \tag{4.5}$$

$$\mathbf{\Sigma} = (\mathbf{D} + \gamma(\mathbf{A}^{\mathbf{n}}_{\mathbf{c}^*})^T \mathbf{A}^n_{c^*})^{-1}, \tag{4.6}$$

where $\mathbf{D}$ is a diagonal matrix with its ith diagonal element is given by $[\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1}]$, for $i = 1, ..., M$. The maximum a posterior (MAP) estimate of $\mathbf{g}'^n_{c^*,c,i}$ is the mean of its posterior distribution, i.e.,

$$\hat{\mathbf{g}}'^n_{c^*,c^*,i} = \boldsymbol{\mu} = ((\mathbf{A}^{\mathbf{n}}_{\mathbf{c}^*})^T \mathbf{A}^n_{c^*} + \gamma^{-1}\boldsymbol{D})^{-1} (\mathbf{A}^{\mathbf{n}}_{\mathbf{c}^*})^T \boldsymbol{y}_{c^*,i}. \tag{4.7}$$

4.2.2 Hyperparameters Estimation

To obtain the term $\hat{\mathbf{g}}'^n_{c^*,c^*,i}$, we need to jointly estimate the hyperparameters $\boldsymbol{\alpha}$ and γ, which can be achieved by exploiting the expectation-maximization (EM) approach (we refer interested readers to [72] and [73] for detailed derivations). So, the new estimate of $\boldsymbol{\alpha}^{(t+1)}$ and $\gamma^{(t+1)}$ can be given as

$$\begin{aligned}\alpha_i^{(t+1)} = 10^{-4}/0.5(\hat{\mu}_i^2 + \hat{\Sigma}_{i,i}) + \beta(\hat{\mu}_{i+1}^2 + \hat{\Sigma}_{i+1,i+1}) \\ + \beta(\hat{\mu}_{i-1}^2 + \hat{\Sigma}_{i-1,i-1}) + 10^{-4},\ i = 1, ..., M,\end{aligned} \tag{4.8}$$

$$\begin{aligned}\gamma^{(t+1)} = M + 2 * 10^{-4}/||\mathbf{y}_{c^*,i} - \mathbf{g}'^n_{c^*,c^*,i}\mathbf{A}^n_{c^*}||_2^2 \\ + (\gamma^{(t)})^{-1} \sum_i (1 - \hat{\Sigma}_{i,i}(\alpha_i^{(t)} + \beta\alpha_{i-1}^{(t)} + \beta\alpha_{i+1})^{(t)}) + 2 * 10^{-4}.\ i = 1, ..., M,\end{aligned} \tag{4.9}$$

The procedures for implementation of the proposed technique are summarized in Algorithms 2.

To verify the merit of the proposed channel estimation technique, the MSE tool is used. The approximated MSE expression for the proposed SBL can be expressed as [52]

Algorithm 2 SBL-based Channel Estimator

INPUTS: Pilot Signal $\boldsymbol{X}^n_{c^*}$, observation matrix $\boldsymbol{y}_{c^*,i}$ and the measurement matrix $\boldsymbol{A}^n_{c^*} = \boldsymbol{F}\boldsymbol{X}^n_{c^*}$

Initial Configuration:

1: Select a specific convergence value $\boldsymbol{\epsilon}$

2: Select a start value for $\boldsymbol{\alpha}^{(t)}$ and $\gamma^{(t)}$

3: $t = 0$

4:**While** $\|(\hat{\mathbf{g}}'^n_{c^*,c^*,i})^{(t+1)} - (\hat{\mathbf{g}}'^n_{c^*,c^*,i})^{(t)}\| \leq \boldsymbol{\epsilon}$ **do**

5: Obtain a new estimate for $\boldsymbol{\alpha}^{(t+1)}$ and $\gamma^{(t+1)}$ as in (15) and (17), respectively.

6: Compute $\boldsymbol{\Sigma} = (\boldsymbol{D} + \gamma(\mathbf{A}^n_{c^*})^T\mathbf{A}^n_{c^*})^{-1}$

7: Compute $\hat{\mathbf{g}}'^n_{c^*,c^*,i} = \boldsymbol{\mu} = \gamma\boldsymbol{\Sigma}\boldsymbol{A}^{n\,T}_{c^*}\boldsymbol{y}_{c^*,i}$

8: $t \leftarrow t + 1$

9: **end**

OUTPUTS: Return the estimated channel $\hat{\mathbf{g}}'^n_{c^*,c^*,i}$

$$MSE = tr\{(\mathbf{D} + \gamma(\mathbf{A}^n_{c^*})^T\mathbf{A}^n_{c^*})^{-1}\}, \tag{4.10}$$

$$= tr\{(\mathbf{D} + \gamma\mathbf{F}^T(\mathbf{X}^n_{c^*})^T\mathbf{X}^n_{c^*}\mathbf{F})^{-1}\}. \tag{4.11}$$

4.3 Optimal Pilot Design

In (4.11), we observe that the MSE of the SBL estimation algorithm relies on the choice of the pilot sequence. Hence, in this section, we design the optimal pilot signals to improve the channel estimation accuracy based on non uniform placements for training signal with respect to MSE of the SBL channel estimator. It has been shown that the minimum MSE can be achieved if $tr((\mathbf{A}^n_{c^*})^T\mathbf{A}^n_{c^*}) = P^{UE}$, where P^{UE} is a fixed power dedicated for training [78]-[80]. To obtain the minimum MSE, we formulate the following optimization problem

To obtain the minimum MSE, we formulate the following optimization problem:

$$\begin{aligned} \underset{X_c^n}{\text{minimize}} \quad & MSE \\ \text{s.t.} \quad & tr((\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n) \leq P^{UE}, 1 \leq n \leq N, \end{aligned} \tag{4.12}$$

where P^{UE} is a fixed power dedicated for training [78]-[80]. Under the power constraint of the MSE estimator, the optimal pilot can be characterized by the following theorem.

The Lagrangian of (4.12) can be formed as follow

$$L(\mathbf{A}_{c^*}^n, \mu) = tr[(\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)^{-1}] + \mu(tr((\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n) - P^{UE}), \tag{4.13}$$

where μ is the Lagrange multiplier. Taking the derivative of (4.13) with respect to $\mathbf{A}_{c^*}^n$ and using the chain rule, we have

$$\begin{aligned} \frac{\partial L(\mathbf{A}_{c^*}^n, \mu)}{\partial(\mathbf{A}_{c^*}^n)} &= \frac{\partial tr[(\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)^{-1}]}{\partial[(\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)]} \\ &\quad \times \frac{\partial tr[\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)]}{\partial \mathbf{A}_{c^*}^n} \\ &\quad + \mu \frac{\partial tr((\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)}{\partial(\mathbf{A}_{c^*}^n)} - \mu \frac{\partial P^{UE}}{\partial(\mathbf{A}_{c^*}^n)}, \end{aligned} \tag{4.14}$$

$$\begin{aligned} &= [-(\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)^{-2}] \\ &\quad \times [\frac{\partial tr(\mathbf{D}^{-1})}{\partial(\mathbf{A}_{c^*}^n)} + \gamma \frac{\partial tr[(\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n]}{\partial(\mathbf{A}_{c^*}^n)}] \\ &\quad + \mu \frac{\partial tr((\mathbf{A}_{c^*}^n)^T \mathbf{A}_{c^*}^n)}{\partial(\mathbf{A}_{c^*}^n)}. \end{aligned} \tag{4.15}$$

Adopting the following identity [81],

$$\frac{\partial tr((\mathbf{Z})^T \mathbf{Z})}{\partial(\mathbf{Z})} = 2\mathbf{Z}, \tag{4.16}$$

one can rewrite (4.15) as follows:

$$\frac{\partial L(\mathbf{A}_{c^*}^n, \mu)}{\partial(\mathbf{A}_{c^*}^n)} = 2[-(\mathbf{D}^{-1} + \gamma(\mathbf{A}_{c^*}^n)^T\mathbf{A}_{c^*}^n)^{-2}]\gamma\mathbf{A}_{c^*}^n + 2\mu\mathbf{A}_{c^*}^n. \tag{4.17}$$

At the optimal point $\frac{\partial L(\mathbf{A}_{c^*}^n, \mu)}{\partial(\mathbf{A}_{c^*}^n)} = 0$. With some simple manipulations, one can write

$$(\mathbf{A}_{c^*}^n)^T\mathbf{A}_{c^*}^n = \gamma^{-1}(\sqrt{\frac{\gamma}{\mu}}\mathbf{I} - \mathbf{D}^{-1}). \tag{4.18}$$

Taking the derivative of (4.13) with respect to μ and equalling it to zero, we have

$$tr\left((\mathbf{A}_{c^*}^n)^T\mathbf{A}_{c^*}^n\right) = P^{UE}. \tag{4.19}$$

Applying trace operation to both sides of (4.18) and using (4.19), we have

$$P^{UE} = tr(\gamma^{-1}(\sqrt{\frac{\gamma}{\mu}}\mathbf{I} - \mathbf{D}^{-1})). \tag{4.20}$$

With some simple manipulations, one can arrive at

$$\sqrt{\frac{\gamma}{\mu}}\mathbf{I} = \gamma(tr(\mathbf{D}^{-1}) + P^{UE})\mathbf{I}. \tag{4.21}$$

Substituting $\sqrt{\frac{\gamma}{\mu}}\mathbf{I}$ by the right hand side of (4.21) into (4.18), the optimal measurement matrix should satisfy the following equation

$$(\mathbf{A}_{c^*}^n)^T\mathbf{A}_{c^*}^n = \gamma^{-1}(\gamma(tr(\mathbf{D}^{-1}) + P^{UE})\mathbf{I} - \mathbf{D}^{-1}). \tag{4.22}$$

Substituting $(\mathbf{A}_{c^*}^n)^T\mathbf{A}_{c^*}^n$ by the right hand side of (4.22) into (4.10), we can obtain the MSE of the proposed estimator in-terms of optimal training design.

The optimal MSE expression for the proposed SBL can be given as

$$MSE = tr\{(\mathbf{D}^{-1} + \gamma(\gamma^{-1}(\gamma(tr(\mathbf{D}^{-1}) + P^{UE})\mathbf{I} - \mathbf{D}^{-1}))^{-1}\}. \tag{4.23}$$

4.4 Achievable Uplink Rate Analysis

Reference (3.4), the base station processes the received signal vector by multiplying it by the uplink maximal ratio combining (MRC) $\boldsymbol{W}_c^H$, as follows

$$\boldsymbol{w}_c^H \mathbf{y}_{c^*,i} = \boldsymbol{w}_c^H \sum_{n=1}^{N} \mathbf{X}_{c^*}^n \mathbf{F} \mathbf{g}'^n_{c^*,c^*,i} + \boldsymbol{w}_c^H \sum_{c=1,c\neq c^*}^{C} \sum_{n=1}^{N} \mathbf{X}_c^n \mathbf{F} \mathbf{g}'^n_{c^*,c,i} + \boldsymbol{w}_c^H \mathbf{v}_{c^*,i}. \tag{4.24}$$

The achievable rate can be expressed as

$$R = E\{log_2(1 + SINR)\}, \tag{4.25}$$

where $SINR$ is the signal to interference and noise ratio

$$SINR = \frac{|\boldsymbol{w}_c^H \sum_{n=1}^{N} \mathbf{F} \mathbf{g}'^n_{c^*,c^*,i}|^2}{|\boldsymbol{w}_c^H \sum_{c=1}^{C} \sum_{n=1}^{N} \mathbf{F} \mathbf{g}'^n_{c^*,c,i}|^2 - |\boldsymbol{w}_c^H \sum_{n=1}^{N} \mathbf{F} \mathbf{g}'^n_{c^*,c^*,i}|^2 + \|\boldsymbol{w}_c\|^2}. \tag{4.26}$$

4.5 CRB For SBL-Based Estimator

To quantify the best performance that can be achieved by the proposed algorithm, in this section, we derive the CRB of the pattern-coupled SBL channel estimation [82]-[84]. The CRB on the covariance of any estimator $\hat{\boldsymbol{\theta}}$ can be given as

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T\} \geq (\mathbf{I} + \mathbf{S}) J^{-1}(\boldsymbol{\theta}) (\mathbf{I} + \mathbf{S})^T, \tag{4.27}$$

where $J(\boldsymbol{\theta})$ is the Fisher information matrix (FIM) that defined in Chapter 3, and $\mathbf{S}$ is the bias gradient matrix defined by that can be given by

$$\mathbf{S} = -\mathbf{I} + \frac{1}{1 + (\mathbf{g}'^{n}_{c^*,c^*,i})^T \mathbf{J} \mathbf{g}'^{n}_{c^*,c^*,i}} \mathbf{g}'^{n}_{c^*,c^*,i} (\mathbf{g}'^{n}_{c^*,c^*,i})^H \mathbf{J}. \tag{4.28}$$

Following the same procedures in section 3.4. The approximated Bayesian CRB for the proposed SBL can be given as

$$J(\mathbf{g}'^{n}_{c^*,c^*,i}) \geq (\frac{1}{(\alpha_i, \beta\alpha_{i+1}, \beta\alpha_{i-1})} + \frac{\mathbf{A}^n_{c^*}(\mathbf{A}^n_{c^*})^T}{\gamma})^{-1}.$$
$$i = 1, ..., M. \tag{4.29}$$

4.6 Simulation Results

In this section, we conduct experiments to evaluate the performance of the proposed algorithm and compare it to existing methods. The simulation parameters used are the same of the previous chapter with the exception of that the scenario is influenced by strong pilot contamination ($\phi^n_{c^*,c^*,i} = 1$ and $\phi^n_{c^*,c,i} = 0.7$). The simulation results are averaged over 1000 channel realizations.

Fig. 4-2 demonstrates the MSE performance comparison of LS with no pilot contamination, SBL-based, thresholded-SBL, Bayesian compressed sensing (BCS) channel estimation techniques along with the BCRB reference line. The results show that the proposed SBL approach provide significant performance enhancement over LS and BCS with respect to estimation accuracy as a result of exploiting the correlation between the antennae and the block sparsity concept.

Fig. 4-3 shows the MSE performance comparison of SBL-based channel estimation with different settings for $\beta = \{0.1, 0.5, \text{ and } 1\}$ for $M = 200$ antennae, $L = 128$ taps and $K = 512$ subcarriers, as indicated earlier in this chapter, β is a

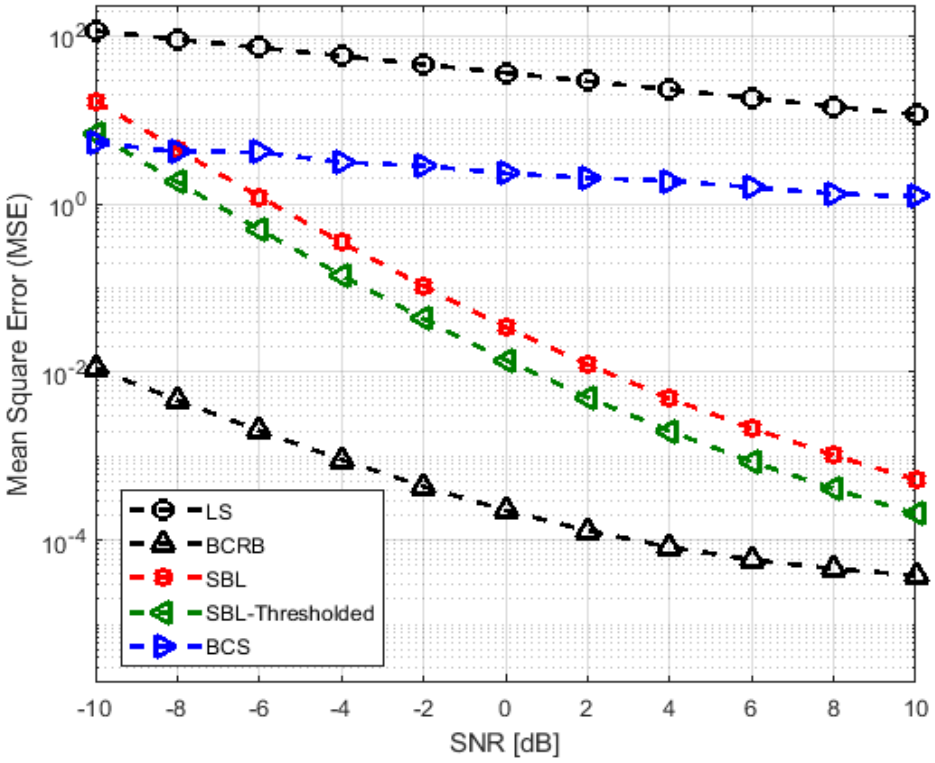

Figure 4-2: Relative MSE performance comparison between SBL, Modified SBL, BCS, and the LS versus SNR.

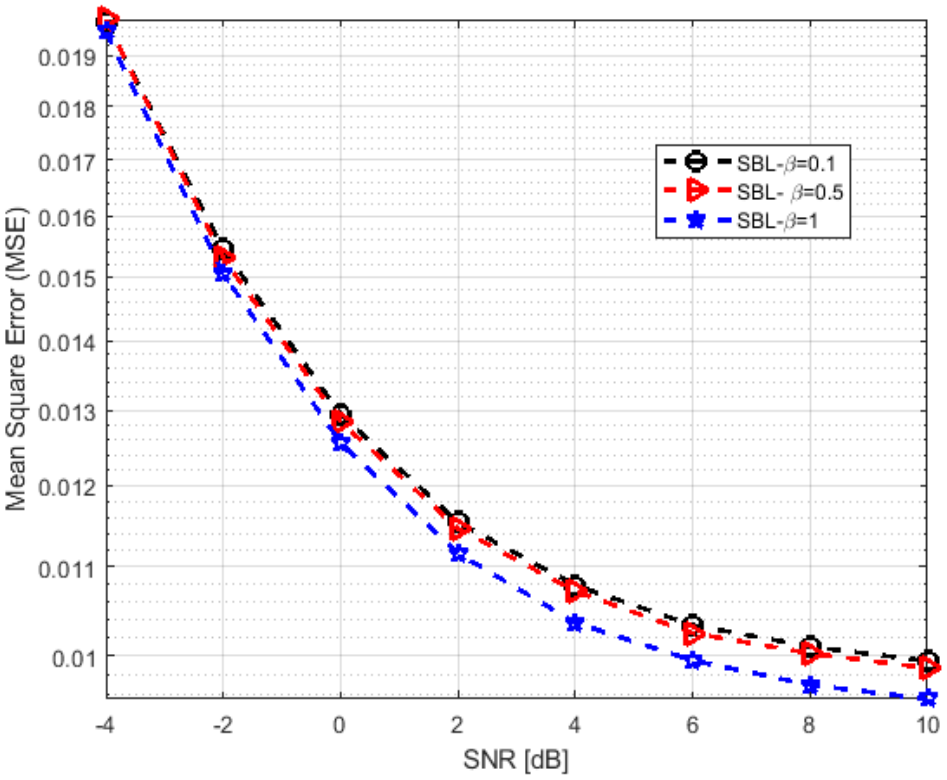

Figure 4-3: Relative MSE of the Pattern-Coupled SBL for $\beta = \{0.1, 0.5 \text{ and } 1\}$.

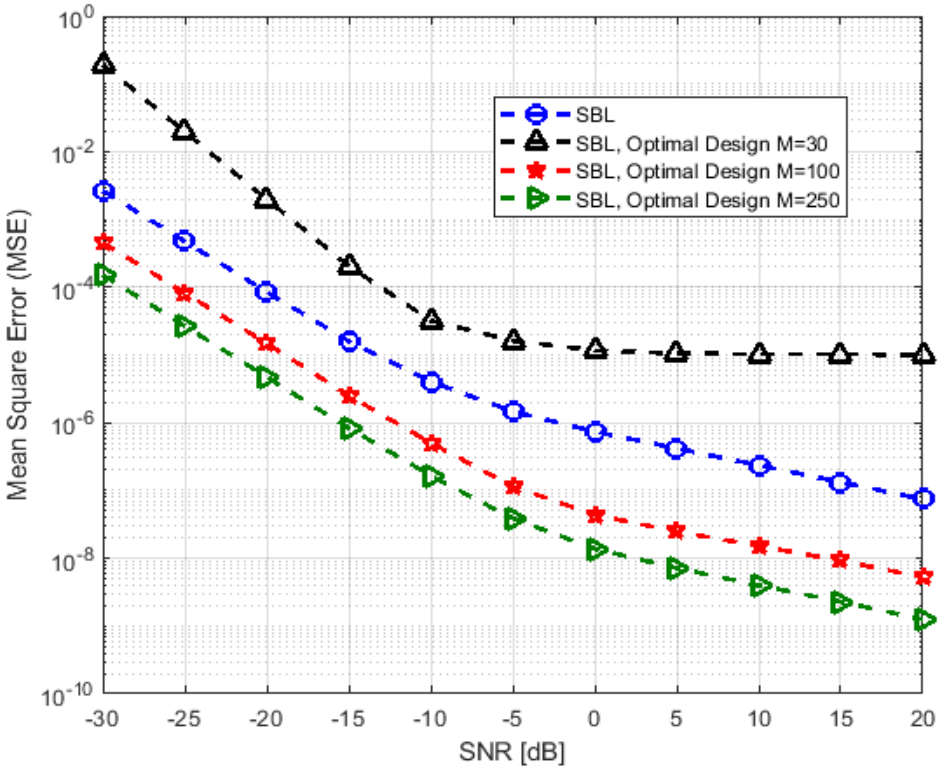

Figure 4-4: Relative MSE of the Pattern-Coupled SBL compared with SBL with optimal pilot scheme for $M = 30, 100$ and 250.

parameter quantifying the dependencies among channel neighbouring coefficients. It can be observed that estimation accuracy is quite comparable to the values of β, as promised by the theory of pattern coupled SBL and that the proposed method is not very sensitive to the choice of β. Thus, it is not required to be either tracked using a little feedback overhead, or estimated using a direct reverse link as the conventional MMSE required to obtain the second order statistics.

Fig. 4-4 shows the performance of the proposed techniques with an optimal pilot design for $\{M = 10, 100$ and $250\}$, $L = 20$ taps and $K = 128$ subcarriers, compared with the conventional SBL-based estimator. The results showed degradation in estimation error for the optimal pilot design based approaches compared with the conventional SBL algorithm.

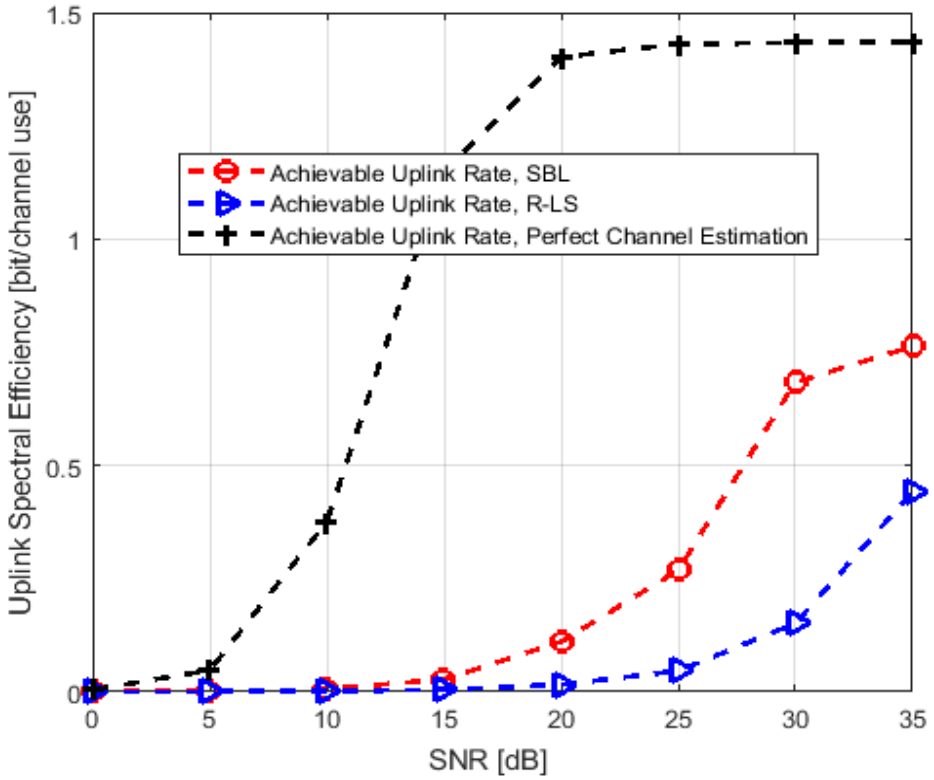

Figure 4-5: Achievable Uplink Rate for R-LS,Pattern Coupled SBL and the perfect channel estimation versus SNR.

Finally, Fig. 4-5 shows the achievable uplink rate achieved by the MRC for LS, SBL estimators compared with the perfect channel estimator. It can be seen that the results demonstrate the previous results. Also it can be observed that with SBL-based estimator we have achieved a comparable performance to the perfect channel estimation. As compared to other receiver, i.e. zero forcing (ZF) and the minimum mean square error (MMSE) receivers, the achievable uplink rate are expected to be enhanced significantly by employing these receivers.

4.7 Conclusion

In this chapter, we proposed a SBL-based channel estimation algorithm for multi-cell massive MIMO systems. The proposed technique jointly exploits prior knowl-

edge on the channel statistic and antenna correlation to provide optimum performance. The simulation results demonstrate that the SBL-based channel estimation algorithm provides better performance over the conventional methods in terms of MSE and the achievable uplink rate. Furthermore, the results demonstrate that the estimation accuracy is enhanced by employing the optimal pilot design sequence set.

Chapter 5

Optimal Pilot Design for Massive MIMO

In this chapter, we consider a multi-cell Massive MIMO system adopting minimum mean square error (MMSE) estimators at the base stations. We derive the mean square error (MSE) of the adopted MMSE estimator as a widely-used criteria of accuracy for estimation. We, then, formulate an optimisation problem to find optimal pilot signals that minimize the total derived MSE of the MMSE estimators of all base stations in the network subject to a transmit power constraint at each user. The proposed formulation is non-convex with respect to the pilot matrices. To overcome non-convexity, we, first, decompose the proposed optimization problem into distributed subproblems at base stations, where each base station in the network optimizes its own pilot signal, given the knowledge of the pilot signals of other base stations. We then introduce a successive optimization approach to transform each subproblem into a linear matrix inequality (LMI) problem which is convex and can be effectively solved by available optimization packages, e.g., [85].

5.1 System Model

Consider a multi-cell massive MIMO system with C cells operating in time division duplexing (TDD) mode. Each cell comprises of M antennae at the base station and N single-antenna users. The propagation factor between the i-th antenna of base station c^* and user n in cell c is $\sqrt{\phi^n_{c,c^*}}h^n_{c,c^*,i}$, where ϕ^n_{c,c^*} is the large scale fading coefficient modeling the path-loss and shadowing, while $h^n_{c,c^*,i} \sim \mathcal{CN}(0,1)$ is small-scale fading. Hereafter, unless otherwise stated, $c \in \{1,2,\cdots,C\}$, $c^* \in \{1,2,\cdots,C\}$, $i \in \{1,2,\cdots,M\}$, and $n \in \{1,2,\cdots,N\}$.

In the pilot training phase, all users in each cell synchronously send their pilot signals. Let $\mathbf{x}^n_c \in \mathbb{C}^{\tau\times 1}$ be the pilot signal used by user n in cell c and

$$\|\mathbf{x}^n_c\|^2 \leq P_{\max,c}, \forall c, \tag{5.1}$$

where τ is the length of the pilot signal, and $P_{\max,c}$ is the maximum allocated power level by each user in cell c to its pilot signal. The received baseband training signal $\mathbf{y}_{c^*,i} \in \mathbb{C}^{\tau\times 1}$ at the i-th antenna element of the base station in cell c^*, denoted as base station c^*, in the network can be expressed as

$$\mathbf{y}_{c^*,i} \;=\; \sum_{c=1}^{C}\sum_{n=1}^{N}\sqrt{\phi^n_{c,c^*}}h^n_{c,c^*,i}\mathbf{x}^n_c + \mathbf{v}_{c^*,i}, \tag{5.2}$$

where $\mathbf{v}_{c^*,i}$ is Gaussian noise with $\mathbf{v}_{c^*,i} \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_\tau)$. Let the received signals, Gaussian noises, pilot signals by all antenna elements of base station c^* and the corresponding large scale channel coefficients be denoted as

$$\mathbf{Y}_{c^*} = [\mathbf{y}_{c^*,1},\ \mathbf{y}_{c^*,2},\ldots,\mathbf{y}_{c^*,M}] \in \mathbb{C}^{\tau\times M}, \tag{5.3}$$

$$\mathbf{V}_{c^*} = [\mathbf{v}_{c^*,1},\ \mathbf{v}_{c^*,2},\ldots,\ \mathbf{v}_{c^*,M}] \in \mathbb{C}^{\tau\times M}, \tag{5.4}$$

$$\mathbf{X}_{c^*} = [\mathbf{x}^1_{c^*},\ \mathbf{x}^2_{c^*},\ldots,\mathbf{x}^N_{c^*}] \in \mathbb{C}^{\tau\times N}, \tag{5.5}$$

$$\mathbf{D}_{c,c^*} = \mathrm{diag}\{[\phi^1_{c,c^*},\ \phi^2_{c,c^*},\ldots,\phi^N_{c,c^*}]^T\} \in \mathbb{C}^{N\times N}. \tag{5.6}$$

Also, let the small-scale fading channel coefficients of all N users in cell c as seen by base station c^* be expressed as

$$\mathbf{H}_{c,c^*} = \begin{bmatrix} h^1_{c,c^*,1}, & \dots, & h^1_{c,c^*,M} \\ \vdots & \ddots & ,\vdots \\ h^N_{c,c^*,1} & \dots & ,h^N_{c,c^*,M} \end{bmatrix} \in \mathbb{C}^{N\times M}. \tag{5.7}$$

Then, using (5.3) − (5.6), one can formulate the received training signals by all M antenna elements of base station c^*, according

$$\mathbf{Y}_{c^*} = \mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c\mathbf{D}^{\frac{1}{2}}_{c,c^*}\mathbf{H}_{c,c^*} + \mathbf{V}_{c^*}. \tag{5.8}$$

The first term in right hand side of (5.8) involves desired channel coefficients and the remaining terms indicate the effects of mutual interference and Gaussian noise. The channel estimate $\widehat{\mathbf{H}}_{c^*,c^*}$ of the original channel $\mathbf{H}_{c^*,c^*}$ is computed by utilizing the MMSE estimation upon the observation of $\mathbf{Y}_{c^*}$ is [86]:

$$\widehat{\mathbf{H}}_{c^*,c^*} = \mathbb{E}[\mathbf{H}_{c^*,c^*}\mathbf{Y}^H_{c^*}]\left(\mathbb{E}[\mathbf{Y}_{c^*}\mathbf{Y}^H_{c^*}]\right)^{-1}\mathbf{Y}_{c^*}. \tag{5.9}$$

Plugging (5.8) in (5.9) and after some mathematical manipulations, we obtain

$$\widehat{\mathbf{H}}_{c^*,c^*} = M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{X}^H_{c^*}\mathbf{\Omega}^{-1}_{c^*}\mathbf{Y}_{c^*}, \tag{5.10}$$

where

$$\mathbf{\Omega}_{c^*} = M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c,c^*}\mathbf{X}^H_c + M\sigma^2\mathbf{I}_\tau. \tag{5.11}$$

See appendix A.

From (5.10), the channel estimation quality depends on the pilot design and if $\tau < CN$, it also suffers from pilot contamination. Let the channel estimation

errors at base station c^* be denoted as

$$\boldsymbol{\Delta}_{c^*} = \widehat{\mathbf{H}}_{c^*,c^*} - \mathbf{H}_{c^*,c^*} \tag{5.12}$$

and the MSE be defined as

$$\mathrm{MSE}_{c^*} = \mathbb{E}\left[\|\boldsymbol{\Delta}_{c^*}\|_F^2\right] = \mathbb{E}\left[\mathrm{tr}\left(\boldsymbol{\Delta}_{c^*}\boldsymbol{\Delta}_{c^*}^H\right)\right]. \tag{5.13}$$

Then, using (5.8), (5.9), and (5.10) and after some mathematical manipulations, one can rewrite MSE in (5.13) as:

$$M\mathrm{tr}\left(\mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}\right), \tag{5.14}$$

where $\mathbf{A}^{-1} = \mathbf{I}_N$, $\mathbf{B} = M\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{X}_{c^*}^H$, $\mathbf{D} = \mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}$, and $\mathbf{C}^{-1} = M\sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}_{c,c^*}\mathbf{X}_c^H + M\sigma^2\mathbf{I}_\tau$. By utilizing the Sherman-Morrison-Woodbury identity [52]

$$(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{D}\mathbf{A}^{-1}$$

and defining $\mathrm{MSE}_{c^*} \triangleq f_{c^*}(\mathbf{X}_{c^*})$, one can reformulate (5.13) as

$$f_{c^*}(\mathbf{X}_{c^*}) = M\mathrm{tr}\left(\left(\mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{X}_{c^*}^H\mathbf{F}_{c^*}^{-1}\mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\right)^{-1}\right), \tag{5.15}$$

where $\mathbf{F}_{c^*} = \sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}_{c,c^*}\mathbf{X}_c^H + \sigma^2\mathbf{I}_N$. See appendix B.

5.2 Optimal Pilot Design

It can be observed from (5.15) that the performance of the MMSE estimation algorithm depends on the pilot structure. Hence, in this section, we develop an optimal pilot design to minimize the total channel estimation errors of all base stations in the network subject to the transmit power constraints at individual

users. To that end, we introduce the following optimization problem for the network as

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} f_{c^*}\left(\mathbf{X}_{c^*}\right) \\ \text{s.t.} \quad & \mathbf{X}_{c^*}^H \mathbf{X}_{c^*} \preceq P_{\max,c^*} \mathbf{I}_N, \forall c^*, \end{aligned} \tag{5.16}$$

where $\{\mathbf{X}_{c^*}\} = \{\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_C\}$. Problem (5.16) is non-convex due to its objective function. To tackle the problem, we first introduce an auxiliary variable $\mathbf{G}_{c^*}$, denote $\{\mathbf{G}_{c^*}\} = \{\mathbf{G}_1, \cdots, \mathbf{G}_C\}$, remove the constant M, and then rewrite (5.16) as

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\},\{\mathbf{G}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} \operatorname{tr}\left(\mathbf{G}_{c^*}\right) \\ \text{s.t.} \quad & \mathbf{X}_{c^*}^H \mathbf{I}_{\tau}^{-1} \mathbf{X}_{c^*} \preceq P_{\max,c^*} \mathbf{I}_N, \forall c^*, \\ & \left(\mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\right)^{-1} \preceq \mathbf{G}_{c^*}, \forall c^*. \end{aligned} \tag{5.17}$$

Adopting the Schur complement [87], one can reformulate problem (5.17) equivalently as

$$\begin{aligned} \underset{\{\mathbf{X}_{c^*}\},\{\mathbf{G}_{c^*}\}}{\text{minimize}} \quad & \sum_{c^*=1}^{C} \operatorname{tr}\left(\mathbf{G}_{c^*}\right) \\ \text{s.t.} \quad & \begin{bmatrix} P_{\max,c^*} \mathbf{I}_N & \mathbf{X}_{c^*}^H \\ \mathbf{X}_{c^*} & \mathbf{I}_{\tau} \end{bmatrix} \succeq \mathbf{0}, \forall c^*, \\ & \begin{bmatrix} \mathbf{G}_{c^*} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0}, \forall c^*. \end{aligned} \tag{5.18}$$

The second set of constraints in (5.18) is still non-convex due to the nonlinearity of the term $\mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}$ with respect to optimization variable $\mathbf{X}_{c^*}$, $\forall c^*$, i.e., the optimization variable is in quadratic forms and appears in both numerator and denominator of the term. As a main contribution of this chapter, we propose a distributed algorithm where every base station c^* optimizes its own pilot signals given the knowledge of the pilot signals of the other cells in $\mathbf{F}_{c^*}^{-1}$, as

follows in the sequel:

$$\begin{aligned} &\underset{\mathbf{X}_{c^*},\mathbf{G}_{c^*}}{\text{minimize}} \quad \text{tr}\left(\mathbf{G}_{c^*}\right) \\ &\text{s.t.} \quad \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^H \\ \mathbf{X}_{c^*} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0}, \\ &\begin{bmatrix} \mathbf{G}_{c^*} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^H \mathbf{F}_{c^*}^{-1} \mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0}. \end{aligned} \tag{5.19}$$

Although (5.19) only considers $\mathbf{X}_{c^*}, \mathbf{G}_{c^*}$ as the optimization variables, its second constraint is still not in an LMI form with respect to $\mathbf{X}_{c^*}$. To proceed, we propose a successive optimization approach where, at the t-th iteration, base station c^* updates its pilot signals by solving the following problem:

$$\begin{aligned} &\underset{\mathbf{X}_{c^*}^{(t)},\mathbf{G}_{c^*}^{(t)}}{\text{minimize}} \quad \text{tr}\left(\mathbf{G}_{c^*}^{(t)}\right) \\ &\text{s.t.} \quad \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^{(t),H} \\ \mathbf{X}_{c^*}^{(t)} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0}, \\ &\begin{bmatrix} \mathbf{G}_{c^*}^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^{(t),H} (\mathbf{F}_{c^*}^{-1})^{(t-1)} \mathbf{X}_{c^*}^{(t-1)} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \end{bmatrix} \succeq \mathbf{0}, \end{aligned} \tag{5.20}$$

where $\mathbf{F}_{c^*}^{-1}$ from the previous iteration is

$$(\mathbf{F}_{c^*}^{-1})^{(t-1)} = \sum_{c=1,c\neq c^*}^{C} \mathbf{X}_c^{(t-1)} \mathbf{D}_{c,c^*} \mathbf{X}_c^{(t-1),H} + \sigma_N^2 \mathbf{I}_N, \tag{5.21}$$

$\mathbf{X}_{c^*}^{(t-1)}$ and $\mathbf{X}_c^{(t-1)}$ are the optimal pilots of cells c^* and c, respectively, which are obtained from the $(t-1)$-th iteration. In order to make the second constraint in (5.20) in LMI form with respect to both $\mathbf{X}_{c^*}^{(t)}$ and $\mathbf{G}_{c^*}^{(t)}$, we have used the known value of $\mathbf{X}_{c^*}^{(t-1)}$. Notice that at the stationary point attained after a sufficient

number of iterations, the approximation

$$\mathbf{X}_{c^*}^{(t)} \approx \mathbf{X}_{c^*}^{(t-1)}, \forall c^*. \tag{5.22}$$

can be assured with any desired accuracy. Note that, still the matrix in the second constraint of (5.20) is not Hermitian during the iterations, due to the mismatch between $\mathbf{X}_{c^*}^{(t)}$ and $\mathbf{X}_{c^*}^{(t-1)}, \forall c^*$. To guarantee a Hermitian matrix in the second constraint of (5.20), we introduce a new variable $\mathbf{A}_{c^*}^{(t)}$, such that

$$\begin{aligned} 2\mathbf{A}_{c^*}^{(t)} =& \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^{(t),H} (\mathbf{F}_{c^*}^{-1})^{(t-1)} \mathbf{X}_{c^*}^{(t-1)} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \\ &+ \mathbf{D}_{c^*,c^*}^{\frac{1}{2}} \mathbf{X}_{c^*}^{(t-1),H} (\mathbf{F}_{c^*}^{-1})^{(t-1)} \mathbf{X}_{c^*}^{(t)} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}. \end{aligned} \tag{5.23}$$

Finally, we reformulate (5.20) as

$$\begin{aligned} \underset{\mathbf{X}_{c^*}^{(t)}, \mathbf{G}_{c^*}^{(t)}, \mathbf{A}_{c^*}^{(t)}}{\text{minimize}} \quad & \text{tr}\left(\mathbf{G}_{c^*}^{(t)}\right) \\ \text{s.t.} \quad & \begin{bmatrix} P_{\max,c^*}\mathbf{I}_N & \mathbf{X}_{c^*}^{(t),H} \\ \mathbf{X}_{c^*}^{(t)} & \mathbf{I}_\tau \end{bmatrix} \succeq \mathbf{0}, \\ & \begin{bmatrix} \mathbf{G}_{c^*}^{(t)} & \mathbf{I}_N \\ \mathbf{I}_N & \mathbf{I}_N + \mathbf{A}_{c^*}^{(t)} \end{bmatrix} \succeq \mathbf{0}, \\ & \text{constraint (5.23)}. \end{aligned} \tag{5.24}$$

Problem (5.24) is now convex and can be efficiently solved by CVX [87]. The procedure to obtain the optimal pilot signals for all C cells in the network is summarized in Algorithm 3.

5.3 Simulation Results

A multi-cell Massive MIMO system is considered for simulations with $C = 4$, $M = 500$, and $N = 10$. All users are randomly distributed over the coverage

Algorithm 3 Successive optimization approach for (5.17)

1: **Inputs:** $\mathbf{D}_{c,c^*}$, $P_{\max,c^*}$, σ^2, stopping criteria $\delta > 0$, initialize $\mathbf{X}_c^{(0)}$, $\forall c, c^*$;
2: $t = 1$;
3: Each cell c^* calculates $\mathbf{F}_{c^*}^{(t-1)}$ utilizing (5.21) and then solves (5.24) to attain $\mathbf{X}_{c^*}^{(t)}$, $\forall c^*$;
4: Exchange $\mathbf{X}_{c^*}^{(t)}$ with the other cells;
5: $\sum_{c^*=1}^{C} \|\mathbf{X}_{c^*}^{(t)} - \mathbf{X}_{c^*}^{(t-1)}\|_F \leq \delta$
6: Go to step 10;
7: $\sum_{c^*=1}^{C} \|\mathbf{X}_{c^*}^{(t)} - \mathbf{X}_{c^*}^{(t-1)}\|_F > \delta$
8: $t = t + 1$;
9: Go to step 3;
10: **Outputs:** $\mathbf{X}_{c^*}^{\star} \leftarrow \mathbf{X}_{c^*}^{(t)}$, $\forall c^*$.

area. However, the distance between any user n of cell c and base station c^*, denoted as d_{c,c^*}^n, is always satisfied $d_{c,c^*}^n \geq 0.035$ km. The system utilizes 20 MHz bandwidth related to the noise variance of -96 dBm and the noise figure of 5 dB. The large-scale fading coefficient $\phi_{c,c*}^n$ [dB] is modeled as

$$\phi_{c,c^*}^n = -148.1 - 37.6 \log_{10}(d_{c,c^*}^n) + z_c^n, \tag{5.25}$$

where z_c^n is the shadow fading which follows a log-normal Gaussian distribution with standard variation 7 dB. Monte-Carlo simulations are tackled over 200 different realizations of user locations. The orthogonal pilot design with an equal power 200 mW assigned to each pilot symbol that is popularly used in many prior works, e.g., [88], [89], is used as a benchmark. The power constraint for pilot signal is set to be $P_{\max,c} = 200\tau$ mW, $\forall c$.

Fig. 5-1 shows the cumulative distribution function (CDF) of the base station-user antenna link which is defined by $f_{c^*}(\mathbf{X}_{c^*})/(MN)$. It is clear from the figure that the channel estimation accuracy of the proposed approach is significantly improved compared to that of the benchmark. This confirms the effectiveness of our optimal pilot design in combating pilot contaminations. The results also indicate that the performance gap between the proposed approach and the benchmark in-

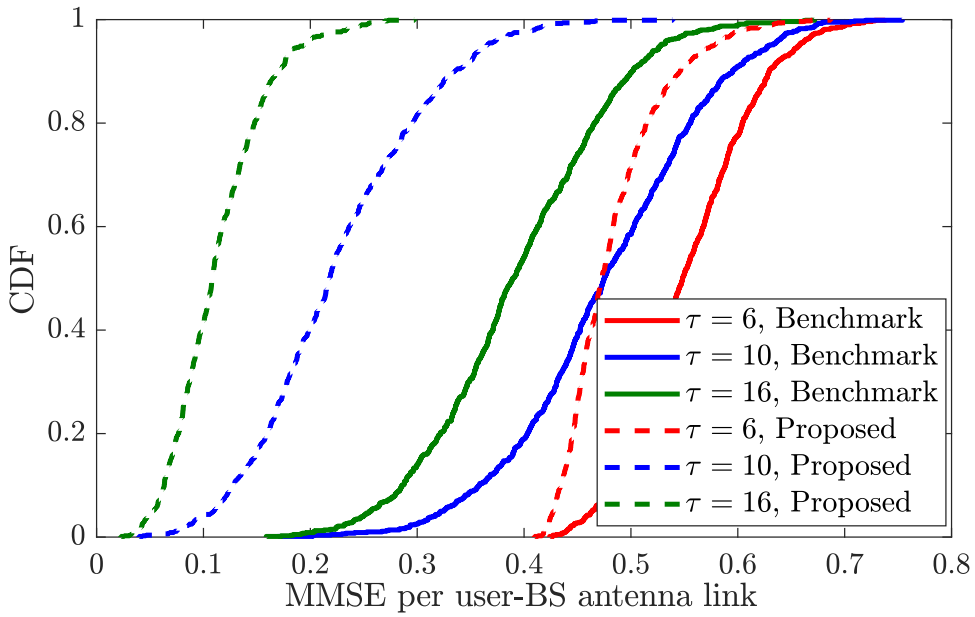

Figure 5-1: The CDFs of the MMSE per user-base station antenna link for the proposed and benchmark approaches in [88] and [89].

creases as the pilot length increases. This is due to a fact that increasing the pilot length gives more degrees of freedom to the proposed approach for optimizing its performances.

Fig. (5-2) (a) displays the CDFs of the power allocated to each pilot symbol with different pilot lengths. It can be seen from the figure that for most of the cases, e.g., around 80 % for $\tau = 6$ and 70 % for $\tau = 10$ and $\tau = 16$, the proposed approach spends less power for each symbol than the benchmark does, i.e., less than 200 mW per symbol.

Fig. (5-2) (b) illustrates the average power for each pilot symbol against the length of pilot signals. When $\tau < N = 10$, the proposed approach consumes less power than the benchmark does. In this case, the number of pilot signals is less than the number of users in each cell. Consequently, the system suffers from both intra-cell and inter-cell interference. In such hostile situations, the proposed approach can still effectively handle the pilot contaminations which results in higher accuracy, i.e., as seen in Fig. 5-1, with less power consumption

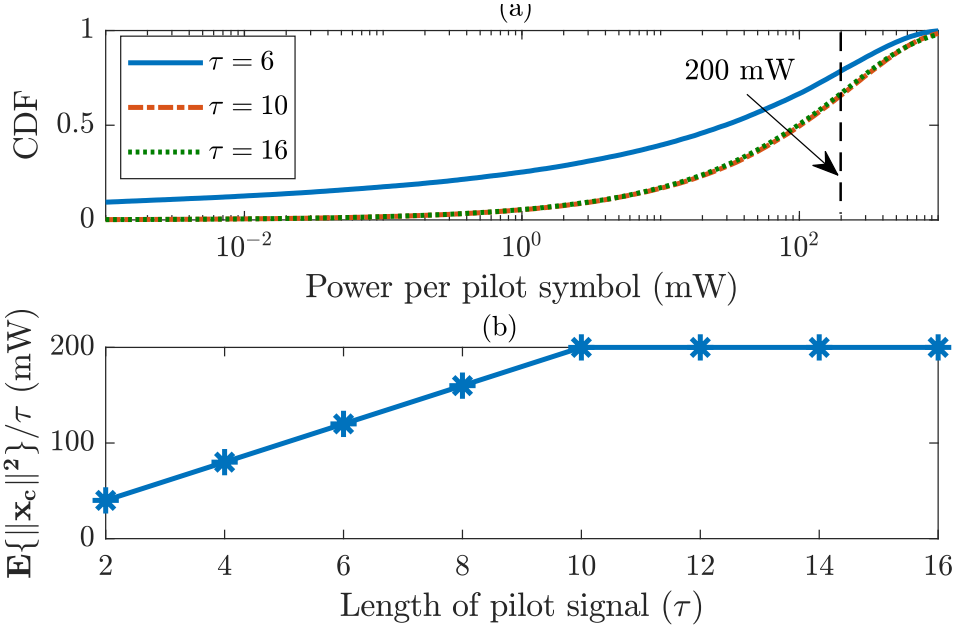

Figure 5-2: (a) The CDFs of the power allocated for each pilot symbol of the proposed approach with different pilot lengths; (b) The average power allocated for each pilot symbol of the proposed approach versus the pilot length.

in comparison with the benchmark. Interestingly, when $\tau \geq N$, the optimal transmit power spent on each symbol turns out to be 200 mW and to be constant irrespective to the value of τ, which can also be observed from the CDF shown in Fig. 5-2 (a).

Finally, Fig.(5.3) reveals the fast convergent speed of the proposed approach, i.e., within less than 20 iterations.

5.4 Conclusion

In this chapter, we tackle the pilot contamination problem in multi-cell Massive MIMO systems. We introduce a novel optimal pilot design approach that minimizes the total mean square errors of the minimum mean square error estimators of all base stations subject to the transmit power constraints of individual users in the network. First, we decompose the original non-convex problem into dis-

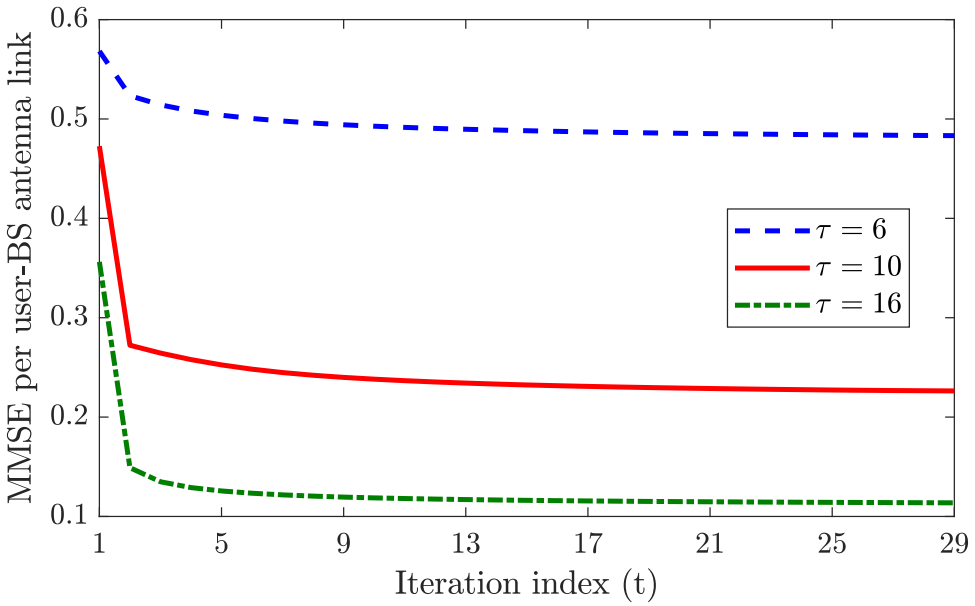

Figure 5-3: The convergence of MMSE per user-base station antenna link versus iteration index for the proposed approach.

tributed optimization sub-problems at individual base stations, where each base station can optimize its own pilot signals given the knowledge of pilot signals from the remaining base stations. We then introduce a successive optimization approach to transform each optimization sub-problem into a linear matrix inequality (LMI) form, which is convex and can be solved by available optimization packages. Simulation results confirm the fast convergence of the proposed approach. The results also indicate that the proposed approach prevails a benchmark scheme in terms of providing higher accuracy.

Chapter 6

Channel Model for Massive MIMO System

In this chapter, we investigate the performance analysis of a multi-cell massive multiple-input multiple-output (MIMO) system that applies a conventional channel estimation technique, namely the minimum mean square error (MMSE) in correlated Rician and correlated Nakagami-m fading channels. Based on this analysis, we found that, by increasing the line-of-sight (LOS) component the channel estimation accuracy is enhanced. Also, we found that the pilot contamination can be eliminated with the usage of very large number of antennae and applying large values of the K-Rice fading factor for the Rician fading channel model. Similarly, the pilot contamination will be eliminated by using large values of antennae and applying large values of the m-shaping factor for the Nakagami-m fading channel. The remainder of this chapter is organized as follows. The multi-cell massive MIMO system model is presented in Section 6.1. The analysis of the achievable uplink spectral efficiency and the pilot contamination analysis are addressed in Section 6.2 and 6.3, respectively. Simulation results are provided in Section 6.4, and the final conclusions are drawn in Section 6.5.

6.1 System Model

Following the system model in [90], we consider a multi-cell massive MIMO with L cells. Each cell consists of M antennae at the base station, N single antennae users, and the system operates in the time-division duplex (TDD) mode to exploit the channel reciprocity. Assuming a block fading structure, each block begins with a uplink pilot, followed by downlink data transmission. The system then toggles to the DL and begins with the uplink data transmission; the coherence period ends the downlink data transmission. The uplink channel is used for pilot-based channel estimation, and the received signal at the base station is expressed as

$$\mathbf{y} = \sqrt{p_u}\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{6.1}$$

where $\mathbf{x}$ is a pilot signal that is used for channel estimation, and the term $\mathbf{n}$ is an ergodic process that consists of independent receiver noise $\mathbf{n}_{noise} \sim \mathcal{CN}(0, \zeta_{BS}^2 I)$ as well as potential interference $\mathbf{n}_{interf}$ from other simultaneous transmissions. We assume that $\mathbf{n}_{interf}$ has zero mean and covariance matrix $\mathbf{S}$ during pilot transmission, where $S = E[\mathbf{n}_{interf}\mathbf{n}_{interf}^H]$, $\mathbf{H}$ is the block of fading of the fast fading matrix between the base station and the user equipment and the average power is $p_u = E[|\mathbf{x}|^2]$ [90]. The fast fading matrix $\mathbf{H}$ is modelled as:

- Correlated Rician fading channel, so $\mathbf{H}_{Rice}$ can be written as [91]:

$$\mathbf{H}_{Rice} = [K(K+1)^{-1}]^{1/2} + \mathbf{H}[(K+1)^{-1}]^{1/2} \tag{6.2}$$

  where K is the Rice factor, which represents the ratio of the power of the deterministic component to the power of the fading component [92].

- Correlated Nakagami-m fading channel, thus $\mathbf{H}_{Nakagami-m}$ can be given as [93]

$$\mathbf{H}_{Nakagami-m} = \mathbf{H}e^{1-m} + \mathbf{H}_{Rice}(1 - e^{1-m}) \tag{6.3}$$

where m is the shaping factor of the Nakagami-m fading that describes the fading degree of the propagation field due to the scattering and multipath interference processes [93].

6.2 Uplink MMSE Channel Estimation

The mean square error (MSE) of the MMSE estimator will be derived in this section. The appropriate relationship of the MMSE estimator for the Rician fading channel that minimises the estimation error of the channel matrix h can be given as in [94] and [95]

$$\hat{\mathbf{H}} = M + (\mathbf{y} - M\mathbf{x})\mathbf{Z}, \tag{6.4}$$

where $\mathbf{Z}$ values are the complex weights chosen to minimise the MSE between the true value of the channel and the estimated channel, and M is the mean value of the channel, and can be given as in [94]

$$M = \sqrt{\frac{K}{K+1}}. \tag{6.5}$$

The MSE of the MMSE estimate of $\mathbf{H}$ can be expressed as

$$MSE = \left\| \mathbf{H} - \hat{\mathbf{H}} \right\|_F^2 \tag{6.6}$$

$$= \left\| \mathbf{H} - M - (\mathbf{y} - M\mathbf{x}) \right\|_F^2, \tag{6.7}$$

$$MSE = tr[E\{(\mathbf{H} - M - (\mathbf{y} - M\mathbf{x})\mathbf{Z})^H (\mathbf{H} - M - (\mathbf{y} - M\mathbf{x})\mathbf{Z})\}], \tag{6.8}$$

Considering:

$$E[\mathbf{y}\mathbf{y}^h] = p_u\mathbf{R} + \mathbf{S} + \zeta_{BS}^2\mathbf{I}, \tag{6.9}$$

Thus, the MMSE estimator can be achieved by differentiating (6.8) with respect to $\mathbf{Z}$ and equating to zero, and the result can be given as

$$\mathbf{Z} = \mathbf{x}^H \mathbf{R}(p_u \mathbf{R} + \mathbf{S} + \zeta_{BS}^2 \mathbf{I})^{-1}. \tag{6.10}$$

Finally, the estimated channel can be expressed as

$$\hat{\mathbf{H}} = M + (\mathbf{y} - M\mathbf{x})\mathbf{x}^H \mathbf{R}(\sqrt{p_u}\mathbf{R} + \mathbf{S} + \zeta_{BS}^2 \mathbf{I})^{-1}. \tag{6.11}$$

6.3 Achievable Spectral Efficiency

Reference the system model in section 6.1, the received vector $\mathbf{y} \in \mathbb{C}^{M\times 1}$ at m-th antenna at the BS can be expressed as

$$\mathbf{y} = \sqrt{p_u}\mathbf{H}_n \mathbf{x}_n + \sqrt{p_u} \sum_{l=1, l\neq n}^{L} \mathbf{H}_l \mathbf{x}_l + \mathbf{n}, \tag{6.12}$$

where $\mathbf{H}_n$ and $\mathbf{x}_n$ are the n-th elements of $\mathbf{H}$ and $\mathbf{x}$, respectively.

The base station processes its received signal vector by multiplying it by the uplink linear MMSE receiver matrix $\mathbf{A}$, as follows:

$$\mathbf{r} = \mathbf{A}^H \mathbf{y}. \tag{6.13}$$

The n-th element of $\mathbf{r}$ can be given as

$$\mathbf{r} = \sqrt{p_u}\mathbf{a}_n^H \mathbf{H}_n \mathbf{x}_n + \sqrt{p_u} \sum_{l=1, l\neq n}^{L} \mathbf{a}_n^H \mathbf{H}_l \mathbf{x}_l + \mathbf{a}_n^H \mathbf{n}, \tag{6.14}$$

where $\mathbf{a}_n$ is the n-th column of $\mathbf{A}$. The achievable rate can be expressed as [96]:

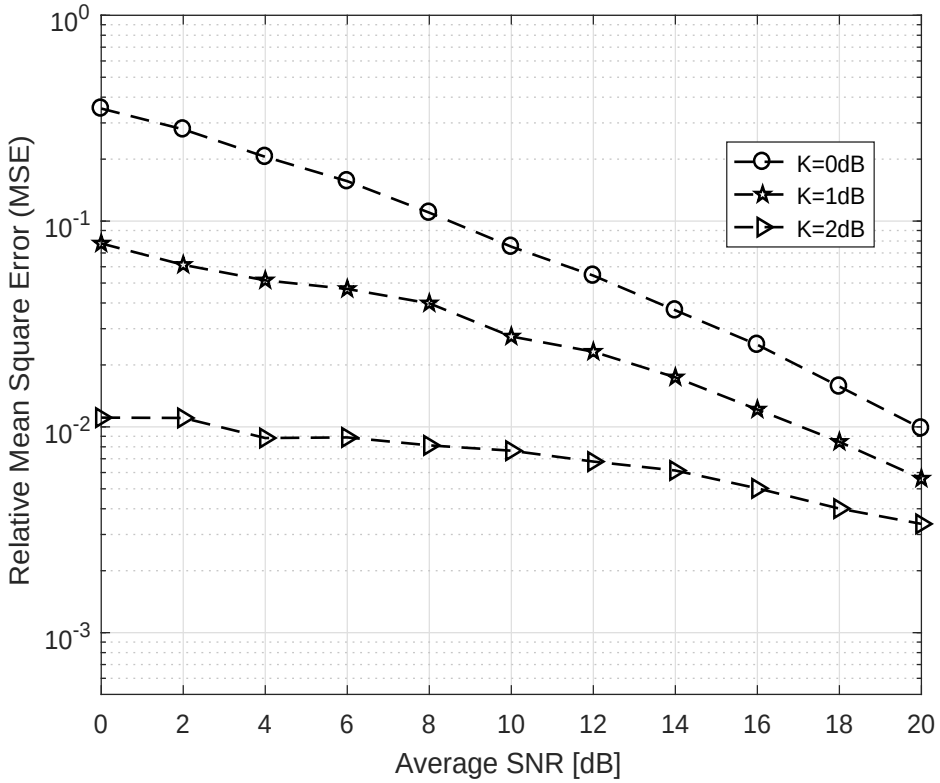

Figure 6-1: The MSE of the MMSE channel estimator versus the average SNR for various Rician K-factors.

$$C = E\{\log_2(1 + SINR)\} \tag{6.15}$$

where $SINR$ is the signal to interference and noise ratio that can be given as

$$SINR = \frac{p_u |E\{\mathbf{a}_n^H \mathbf{H}_n\}|^2}{p_u \sum_{l=1}^{L} E\{|\mathbf{a}_n^H \mathbf{H}_n|^2\} - p_u |E\{\mathbf{a}_n^H \mathbf{H}_n\}|^2 + \|\mathbf{a}_n\|} \tag{6.16}$$

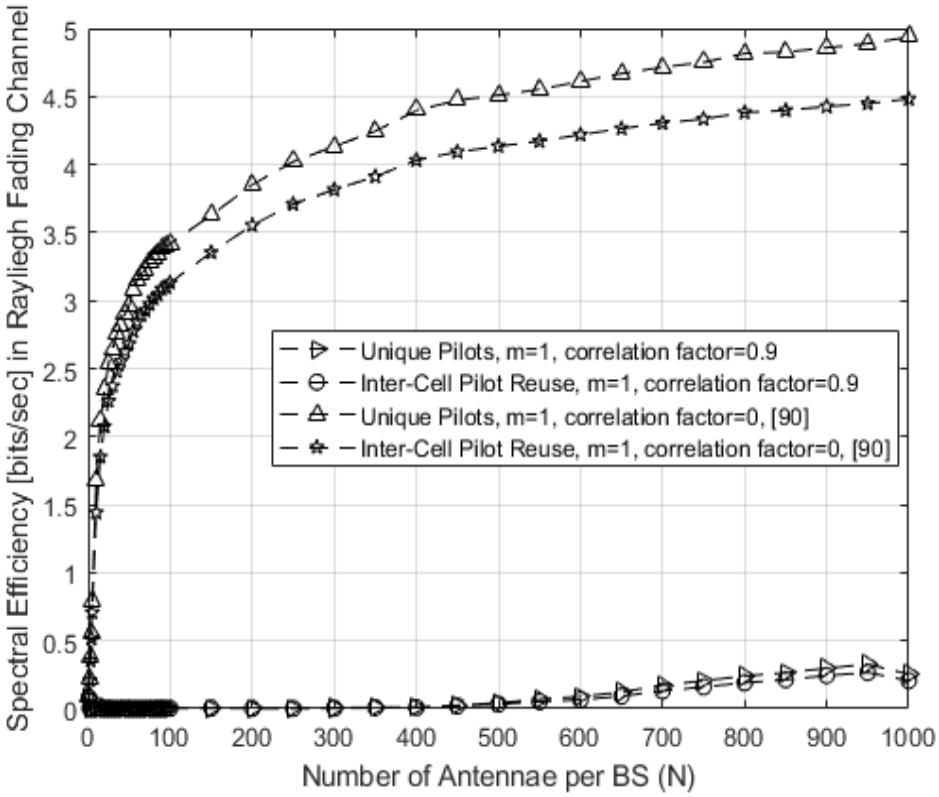

Figure 6-2: The achievable uplink spectral efficiency of the multi-cell multi-user massive MIMO in correlated Rayleigh fading versus the results reproduced from [90].

6.4 Numerical Results

In this section, the MSE results over correlated Rician fading channels are provided. The relative MSE is to be normalised with the channel covariance matrix, and can be written as

$$MSE = \frac{\left\| \mathbf{H} - \hat{\mathbf{H}} \right\|_F^2}{tr(\mathbf{R})} \tag{6.17}$$

The channel covariance matrix $\mathbf{R}$ is generated by the exponential correlation model [97]. The system model considers the effect of large scale fading in terms of path losses and the shadowing effect of the LTE-system model, as in the 3GPP

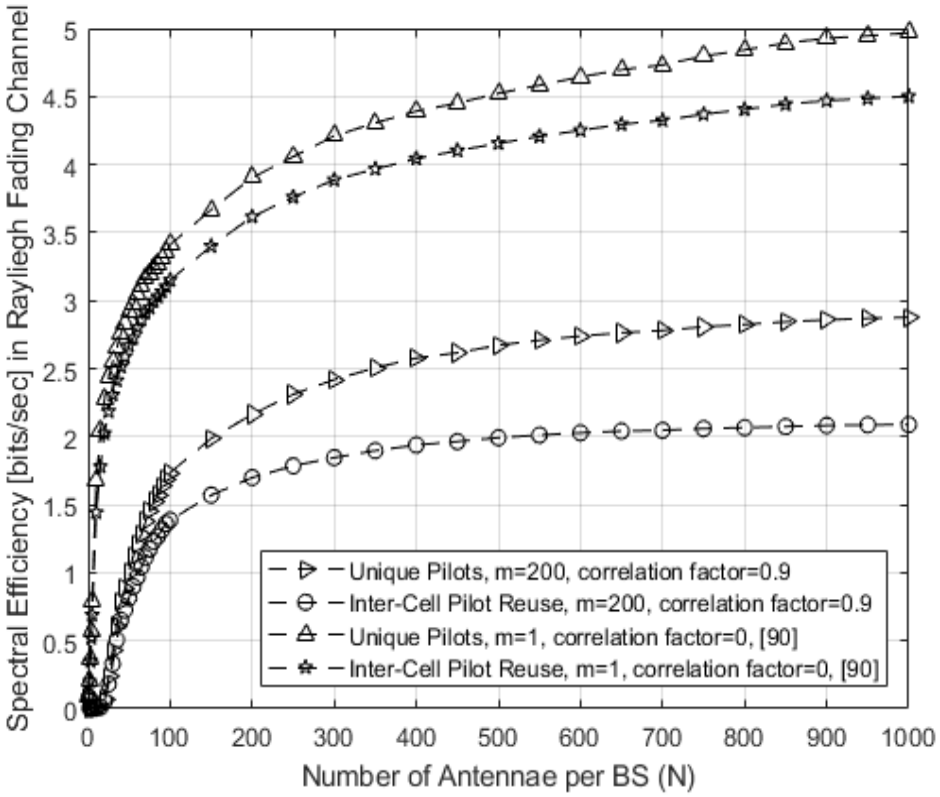

Figure 6-3: The achievable uplink spectral efficiency of multi-cell multi-user massive MIMO in correlated Nakagami-m fading versus the results reproduced from [90].

propagation model in [98]. We assume the number of users is 10, the number of base station antennae is 100 and the correlation factor is 0.7. In Fig.6-1, the MSE versus the signal-to-noise ratio (SNR) in dB in the presence of the pilot contamination for K-Rice factor = 1,10 and 100 is plotted, and it can be clearly observed that the estimation accuracy of the MMSE is enhanced as the line-of-sight (LOS) component increases via increasing the values of K-Rice fading factor. Fig.6-2 illustrates the achievable uplink spectral efficiency of a multi-cell massive MIMO for both the scenarios of unique and reused pilots in highly correlated Rayleigh fading with a correlation factor of 0.9 versus the results reproduced form [90]. It can be seen the achievable spectral efficiency is dropped by the effect of high correlation.

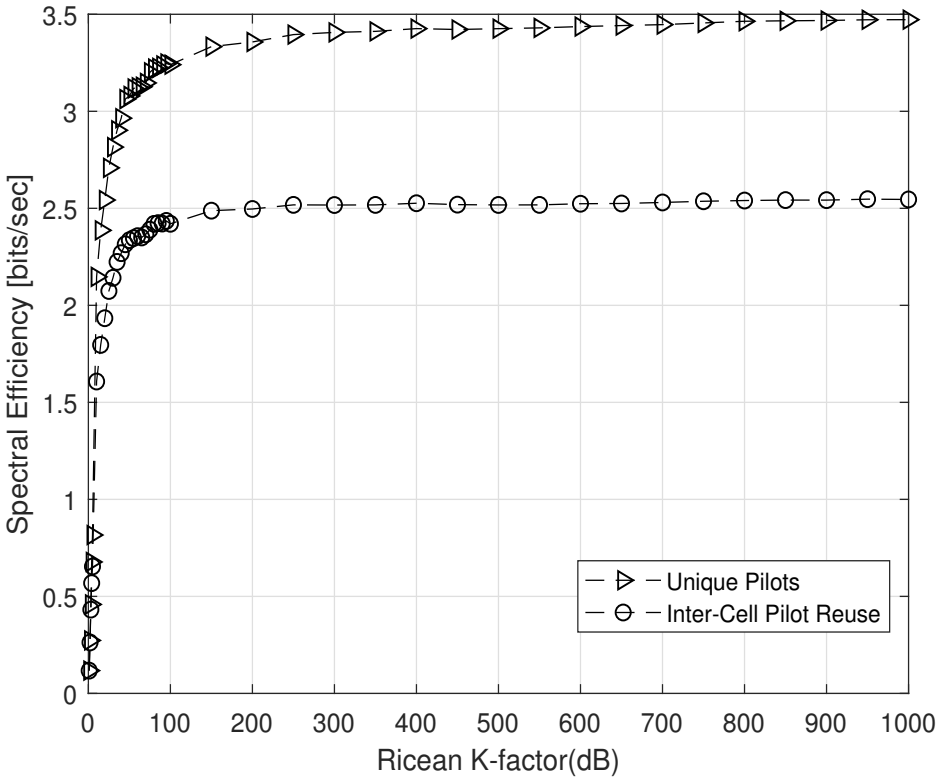

Figure 6-4: The achievable uplink spectral efficiency of multi-cell multi-user massive MIMO over correlated Rician fading versus the K-Rice fading factor for the unique pilot and reused pilot scenarios.

Fig. 6-3 illustrates the achievable uplink spectral efficiency of the multi-cell massive MIMO for both scenarios of unique and reused pilots in correlated Nakagami-m fading with a correlation factor of 0.9 and an m-shaping factor of about 200, versus the results reproduced from [90]. It can be seen that the effect of the correlation can be reduced with a high value of m-shaping factor. Fig. 6-4 and Fig. 6-5 illustrate the achievable uplink spectral efficiency of the massive MIMO versus the K-Rice factor and the m-shaping factor, respectively, it can clearly be observed that the effect of pilot contamination can be eliminated.

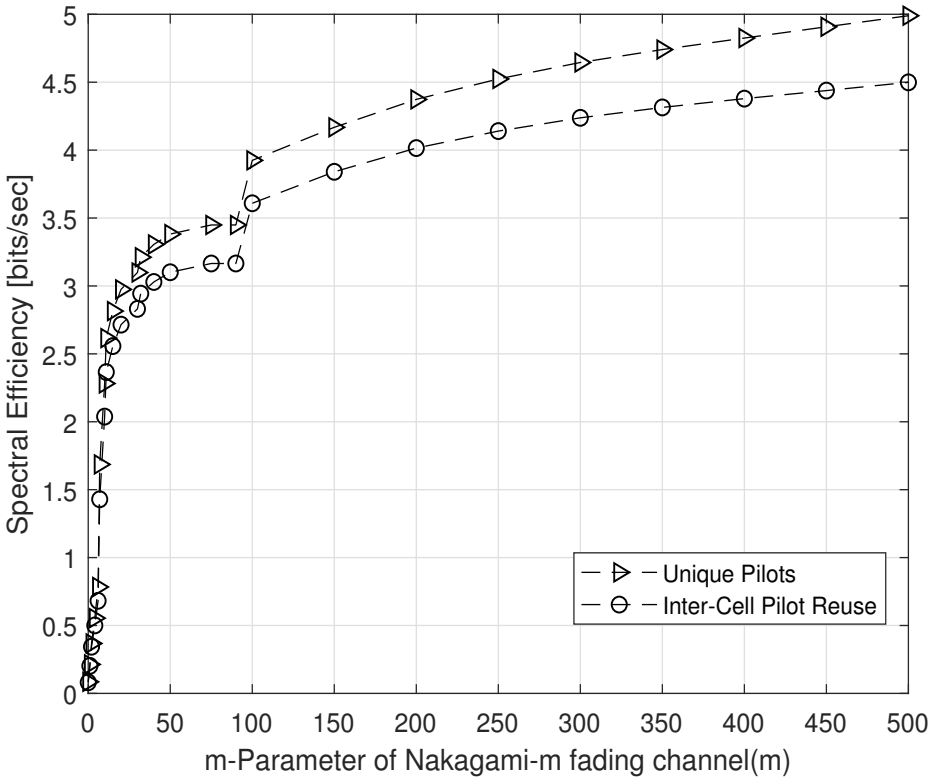

Figure 6-5: The achievable uplink spectral efficiency of the multi-cell multi-user massive MIMO over correlated Rician fading versus the m-shaping factor for the unique pilot and reused pilot scenarios.

6.5 Conclusion

In this chapter, the performance analysis of a massive MIMO in a multi-cell scenario over correlated Rician fading and correlated Nakagami-m fading channel models have been investigated. It can be found that, as the values of the K-Rice factor increase, the estimation accuracy in terms of the MSE of the MMSE is enhanced. Furthermore, based on the achievable uplink spectral efficiency results, the pilot contamination problem can be eliminated when applying large values for the K-Rice factor and using a large number of antennae. The same sort of results of the elimination of the pilot contamination can be achieved for correlated Nakagami-m fading when employing a very large number of antennae and large

values for the m-shaping factor. In addition, it can be observed that, for very large values of the K-Rice factor or the m-shaping factor of the Nakagami-m, the achievable uplink spectral efficiency approaches a constant value. The results may assist the system designer to determine how the system and channel parameters affect the reliability of the achievable spectral efficiency and MSE of massive MIMO systems. The theoretical simulation and closed form approximation for the MSE and the achievable rate for the correlated Rician fading and the correlated Nakagami-m fading is left for future research and as an extension to the work in this research.

Chapter 7

DFT-Based Channel Estimation Techniques

In this chapter, the pilot contamination problem of massive MIMO systems is addressed using three modified DFT-based channel estimation approaches, namely, the iterative DFT, the DFT-based Most Significant Taps Approach (DMSTA) and the joint DFT with whitening rotation (WR) semi blind estimation.

In DFT-based approach, the channel frequency response (CFR) is first estimated by applying the conventional channel estimation i.e. linear minimum mean-square error (LMMSE) algorithm in the frequency domain. Then, the IDFT is applied to obtain the channel impulse response (CIR) in time domain [99], [100]. Thereafter, a truncation of the CIR is performed to improve the estimation accuracy in the time domain. Finally, DFT is applied to the truncated CIR to obtain the estimated CFR [76], [77]. In iterative DFT, the procedure in DFT estimation described previously is executed several times to address the pilot contamination problem. While in DMSTA, more CIR truncation will be applied based on threshold value to select the taps with the most significant power [99], [101] and [102].

As explained in the previous chapters, pilot based estimation for massive

MIMO systems suffers from pilot contamination, while the blind based estimation suffers from high computationally complexity. Thus, in the last part of this chapter, we propose a channel estimation scheme that employ a limited number of pilots to address the pilot contamination problem and to alleviate the complexity of the blind method. Such a scheme is known as a semi-blind, since it employs both pilot and blind information [6]. The WR technique consists of two steps: 1) estimation of the whitening matrix using information data; and 2) estimation of the unitary matrix using pilots [67]. We combine the DFT-based channel estimator presented in the previous section with the WR technique. The results show that the proposed scheme provides better performance compared to the WR based and the cramer-rao bound (CRB) reference line.

The remainder of this chapter is organised as follows. The massive MIMO system model is described in Section 7.1. The conventional-DFT based estimator is presented in Section 7.2. The iterative DFT-based channel estimation is introduced in Section 7.3. The DMSTA-Based is presnsted in section 7.4. WR-Based semi-blind estimator is addressed in Section 7.5. Numerical results are provided in Section 7.6. Finally, conclusions are drawn in Section 7.7.

7.1 Massive MIMO System Model

The system model considered in this chapter is the same as the one in Chapter 3. Thus reference (3.1), the estimated channel using the conventional LMMSE estimator that completed entirely in the frequency domain based upon the observation of $Y_{c^*,i}[k]$ can be expressed as

$$\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k] = E[h^n_{c^*,c^*,i}[k](y_{c^*,i}[k])^H](E[y_{c^*,i}[k] \\ (y_{c^*,i}[k])^H])^{-1} y_{c^*,i}[k], \tag{7.1}$$

$$\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k] = E[h^n_{c^*,c^*,i}[k](\sum_{n=1}^{N} h^n_{c^*,c^*,i}[k]x^n_{c^*}[k]+ \sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N} h^n_{c^*,c,i}[k]x^n_c[k] + v_{c^*,i}[k])^H] E[(\sum_{n=1}^{N} h^n_{c^*,c^*,i}[k]x^n_{c^*}[k] + \sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N} h^n_{c^*,c,i}[k] x^n_c[k] + v_{c^*,i}[k])(\sum_{n=1}^{N} h^n_{c^*,c^*,i}[k]x^n_{c^*}[k]+ \sum_{c=1,c\neq c^*}^{C}\sum_{n=1}^{N} h^n_{c^*,c,i}[k]x^n_c[k] + v_{c^*,i}[k])^H])^{-1} y_{c^*,i}[k]. \quad (7.2)$$

Assuming that $h^n_{c^*,c^*,i}[k]$ consists of independent, complex Gaussian components with zero mean and unit variance, then $E\{h^n_{c^*,c^*,i}[k](h^n_{c^*,c^*,i}[k])^H\} = I$. Hence (7.3) can expressed as

$$\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k] = \sum_{n=1}^{N}(x^n_{c^*}[k])^H \underbrace{[\sum_{n=1}^{N} x^n_{c^*}[k](x^n_{c^*}[k])^H + \sum_{c=1}^{C}\sum_{n=1}^{N} x^n_c[k](x^n_c[k])^H + \sigma_v^2 I]^{-1}}_{=\xi} y_{c^*,i}[k]. \quad (7.3)$$

7.2 DFT-Based Channel Estimator

To improve the LMMSE estimation accuracy, the DFT-based channel estimation algorithm has been proposed to reduce the noise and interference components in the time domain. The LMMSE estimated sample of CIR can be expressed as

$$\underline{h}^{n,LMMSE}_{c^*,c^*,i}[\omega] = IDFT_k\{\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k]\}, \quad (7.4)$$

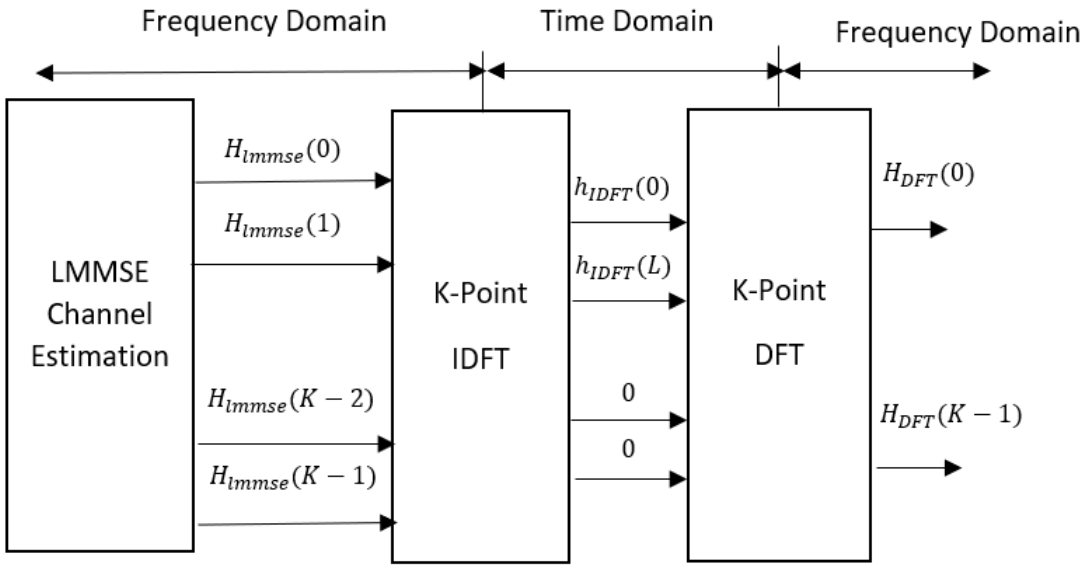

Figure 7-1: DFT-based channel estimation [99].

for $1 \leq k \leq K$, where $IDFT_k\{\}$ indicates K-point IDFT and ω is the time domain sample index. As the estimated CIR from the LMMSE has most of its power concentrated in a few initial samples, the CIR is typically limited to the number of the channel taps L, which is much smaller than the number of subcarriers. Hence, (7.5) can be expressed as:

$$\hat{\underline{h}}_{c^*,c^*,i}^{n,IDFT}[\omega] = \begin{cases} \underline{h}_{c^*,c^*,i}^{n,LMMSE}[\omega], \ 0 \leq l \leq L \\ 0, \ Others \end{cases} . \tag{7.5}$$

In doing so, more noise and interference are cancelled, and the intended channel information is retained. Next, the DFT operation is conducted to recover the channel responses into the frequency domain, as follow:

$$\hat{h}_{c^*,c^*,i}^{n,DFT}[k] = DFT_k\{\hat{\underline{h}}_{c^*,c^*,i}^{n,IDFT}[\omega]\}. \tag{7.6}$$

From what has been described earlier, it is clear that compared with LMMSE, the DFT based channel estimation method makes use of IDFT/DFT to suppress

noise and interference in the time domain [20], as shown in the basic block diagram of the DFT-based estimation in Fig. 7-1 (for notational simplicity in Fig. 7-1, we refer to $\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k], \hat{h}^{n,IDFT}_{c^*,c^*,i}[w]$ and $\hat{h}^{n,DFT}_{c^*,c^*,i}[w]$ as $H_{lmmse}(k), h_{IDFT}(L)$ and $H_{DFT}(k)$, respectively).

We now derive an expression for the approximated $\hat{h}^{n,DFT}_{c^*,c^*,i}[k]$, by taking the IFFT and FFT to the CFR of the LMMSE estimated channel multiplied by the rectangular function that represent the limitation of the CIR of the DFT-based estimator definition that can be written as follow

$$\hat{h}^{n,DFT}_{c^*,c^*,i}[k] = \mathcal{F}\{Rect[\omega]\mathcal{F}^{-1}\{\hat{h}^{n,LMMSE}_{c^*,c^*,i}[k]\}\}, \tag{7.7}$$

where $\mathcal{F}/\mathcal{F}^{-1}$ stands for the IDFT/DFT operations, respectively, and $Rect[\omega]$ can be given as:

$$Rect[\omega] = \begin{cases} 1, & \omega < L \\ 0, & \omega \geq L \end{cases},$$

$$\hat{h}^{n,DFT}_{c^*,c^*,i}[k] = \{\{Rect[\omega]\} \circledast \hat{h}^{n,LMMSE}_{c^*,c^*,i}[k], \tag{7.8}$$

$$\hat{h}^{n,DFT}_{c^*,c^*,i}[k] = \psi(k) \circledast \hat{h}^{n,LMMSE}_{c^*,c^*,i}[k], \tag{7.9}$$

where $\boldsymbol{\psi}(k)$ can be given as
$\boldsymbol{\psi}(k) = \frac{1}{\sqrt{LK}} e^{\frac{-j\pi(L-1)}{K}} Sinc(\frac{\pi kL}{K})$.

7.3 Iterative DFT-Based Channel Estimation

To address the pilot contamination problem, we proposed an iterative algorithm for a DFT-based estimator for a massive MIMO-OFDM system. Accurate channel estimation can be obtained by applying algorithm 4 [102], [103].

Algorithm 4 Iterative-based DFT Channel Estimation

1: **Inputs:** Number of the channel taps L, Threshold value δ.
2: Perform the LMMSE estimation to obtain the $\hat{h}_{c^*,c^*,i}^{n,LMMSE}[k]$.
3: $p = 0$
4: Perform the IFFT to transform the CFR to time domain, as described in section III.
5: Truncate the CIR for the time delay $L-1$.
6: Perform the DFT-based estimator to obtain $\hat{h}_{c^*,c^*,i}^{n,DFT,p}[k]$.
7: **Iterative Procedure:**
8: If $\delta \leq max|\hat{h}_{c^*,c^*,i}^{n,LMMSE,p}[k] - \hat{h}_{c^*,c^*,i}^{n,LMMSE}[k]|$, then $p = p+1$ and **Repeat:** 4 to 6.
9: **Else if:**
10: **Output:** Return the estimated channel $\hat{h}_{c^*,c^*,i}^{n,DFT,p+1}[k]$

7.4 DMSTA-Based Approach

Practically, the CIR contains many taps with no significant energy, by neglecting those taps, the noise and interference through pilot contamination will be eliminated and that can improve the channel estimation performance significantly. In this section, we will use thresholding approach by retaining the channel taps whose energy is above a threshold value η and set the other taps to zero, while the suitable value of η will be determined based on the facts mentioned earlier in this chapter, that most channel taps will concentrated on the lower region in time domain while the noise and pilot contamination components will be spread over the CIR. So, by estimating the power of the noise and interference by averaging the samples on the noise and interference region [104]-[106], the threshold value can be given as follows

$$\eta = \zeta\alpha. \tag{7.10}$$

where ζ can be given as

$$\zeta = \frac{1}{L}\sum_{k=L}^{K} |\hat{h}_{c^*,c^*,i}^{n,DFT}|^2. \tag{7.11}$$

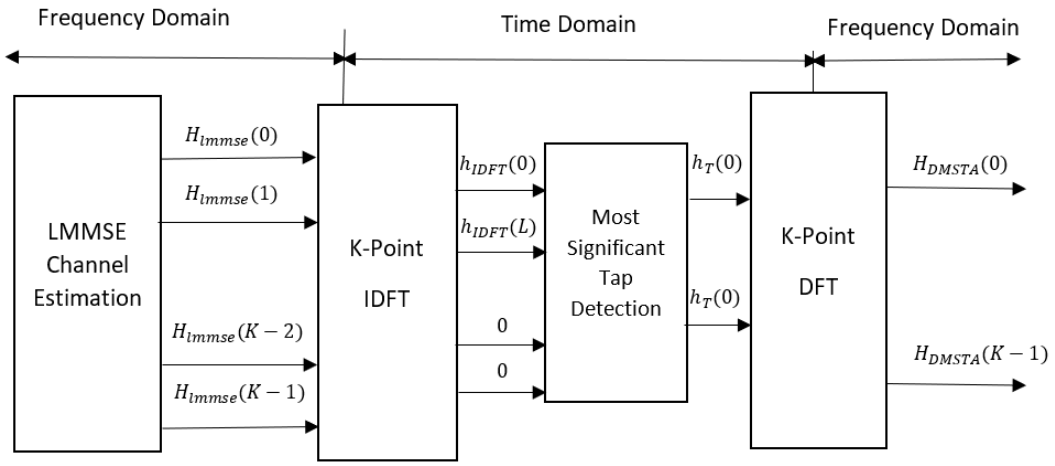

Figure 7-2: DMSTA-based channel estimation [99].

and α is a scaling factor that can be adjusted as noise margin.

So, the most significant taps block will detect the significant channel taps [104]-[107], as follows:

$$\hat{h}^{n,T}_{c^*,c^*,i} = \begin{cases} \hat{h}^{n,DFT}_{c^*,c^*,i}, |\hat{h}^{n,DFT}_{c^*,c^*,i}| > \eta \\ 0, \ otherwise \end{cases} . \tag{7.12}$$

Therefore, the DMSTA-based channel estimation algorithm can be expressed in the frequency domain as

$$\hat{h}^{n,DMSTA}_{c^*,c^*,i} = DFT\{\hat{h}^{n,T}_{c^*,c^*,i}\} \tag{7.13}$$

More elaborations can be found in Fig. 7-2.

7.5 WR-Based Semi-Blind Channel Estimation Algorithm

Based on the WR algorithm that proposed and analysed in [67], [110]-[112], the channel matrix $\mathbf{H}$ can be decomposed as

$$\boldsymbol{H} = \mathcal{W}\mathbf{Q}^H, \tag{7.14}$$

where $\mathcal{W}$ is a whitening matrix and $\mathbf{Q}$ is a unitary rotation matrix.

By performing the singular value decomposition (SVD) to the $\boldsymbol{H}$, we can obtain

$$\boldsymbol{H} = \mathbf{U}_W \boldsymbol{\Xi}_W \mathbf{V}_W^H, \tag{7.15}$$

So, one possible solution of $\mathcal{W}$ and $\mathbf{Q}^H$ can be $\boldsymbol{U}_W\boldsymbol{\Xi}_W$ and $\mathbf{V}_W$, respectively. Now, the concern is how to estimate the whitening matrix $\mathcal{W}$ and rotation matrix $\mathbf{Q}$. The whitening matrix can be estimated blindly from using the second order statistic of the received data. Thus, we can assume that the whitening matrix is perfectly known at the base station.

The unitary matrix $\boldsymbol{Q}$ can be estimated from the pilot symbol sequence $\boldsymbol{X}_{c^*}$ by minimizing the following optimization problem [76], [110]-[112].

$$\underset{\{\mathbf{Q}\}}{\text{minimize}} \|\boldsymbol{Y}_{c^*} - \mathcal{W}\boldsymbol{Q}^H\boldsymbol{X}_{c^*}\|^2, \tag{7.16}$$

$$\text{s.t. } \boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{I}_N \tag{7.17}$$

By letting

$$\boldsymbol{M} = \mathcal{W}^H\boldsymbol{Y}_{c^*}\boldsymbol{X}_{c^*}^H, \tag{7.18}$$

the rotation matrix $\boldsymbol{Q}$ can be calculated by

$$\hat{\boldsymbol{Q}} = \boldsymbol{V}_Q \boldsymbol{U}_Q^H, \tag{7.19}$$

where $\boldsymbol{U}_Q$ and $\boldsymbol{V}_Q$ are obtained from an SVD of M as

$$\boldsymbol{M} = \boldsymbol{U}_Q \boldsymbol{\Xi}_Q \boldsymbol{V}_Q^H. \tag{7.20}$$

By estimating $\mathcal{W}$ and $\boldsymbol{Q}$ matrices, the channel matrix $\boldsymbol{H}$ can be given from (8.5) [76], [110]-[112].

7.6 Numerical Results

In this section, various computer simulations are carried out to evaluate the performance of the LMMSE, DFT-based and iterative-DFT estimators. We assumed that $N = 20$ users, $M = 100$ antennae and the system under the influence of strong pilot contamination i.e. $\phi_{c^*,c^*,i} = 1$ and $\phi_{c^*,c^*,i} = 0.7$. The relative MSE can be written as

$$MSE = \frac{E\{||h_{c^*,c^*,i}^{n,DFT} - \hat{h}_{c^*,c^*,i}^{n,DFT}||_F^2\}}{E\{||h_{c^*,c^*,i}^{n,DFT}||_F^2\}}. \tag{7.21}$$

Fig. 7-3 shows the MSE performance of the DFT-based channel estimation with different values of $K = \{256, 512 \text{ and } 1024\}$, $L = 128$ taps, so the compression ratio (CR), i.e. L/K is to be $CR = \{0.5, 0.25 \text{ and } 0.125\}$. The results indicate that the performance of the conventional DFT-based estimator is enhanced when decreasing the number of the CR, as a result of eliminating more components of noise and interference.

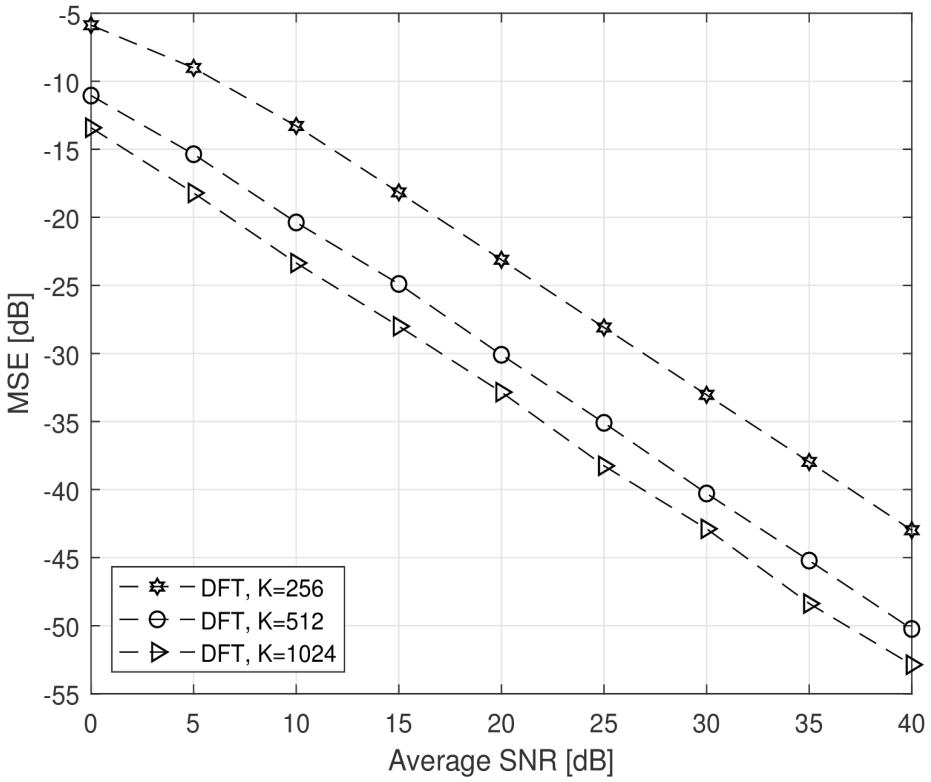

Figure 7-3: The MSE of the DFT-based channel estimation versus the SNR for the number of path is 128 and for different values of $\{K = 256, 512 \text{ and } 1024\}$ so the compression ratio (CR) (i.e. L/K) is to be $CR = \{0.5, 0.25 \text{ and } 0.125\}$.

Fig. 7-4 demonstrates the MSE performance of the conventional LMMSE, DFT, the modified DFT-based estimator in [107], and the iterative DFT-based channel estimation with one, two and three iterations. The results obviously show the iterative DFT estimators outperforms the other estimators.

Fig. 7-5 demonstrates the MSE performance of the conventional LMMSE, DFT, DMSTA, the modified DFT-based estimator in [107]. Obviously, it can be seen that the DMSTA provide better estimation performance compared to other estimators.

Fig. 7-6 and Fig. 7-7 shows the uplink achievable rate of the matched receiver combining (MRC) for the iterative DFT-based (2-5 iterations) and DMSTA, respectively. The performance of the proposed techniques is compared with

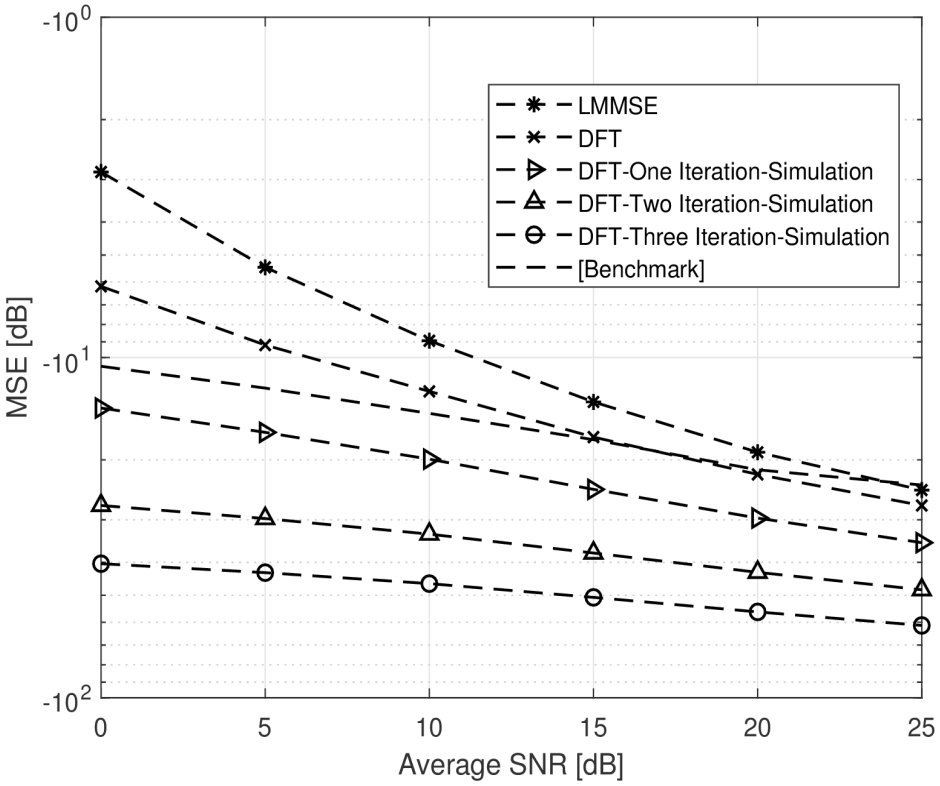

Figure 7-4: The MSE comparison of LMMSE, DFT and iterative-DFT-based estimators as a function of SNR.

the LMMSE with pilot contamination, conventional DFT-based,and the exact LMMSE (LMMSE with no noise and pilot contamination). It can be seen that the iterative DFT-based estimators and DMSTA performed closed to the exact LMMSE estimator with increasing the iterations. This demonstrates a significant improvement in addressing the pilot contamination problem.

From the results in Fig. 7-8, the MSE has been plotted versus SNR in the range of 0dB to 35dB for the LMMSE, WR, DFT and the joint DFT and WR for (L=10, 50 and 100, respectively). The CRB of the semi-blind scheme is also plotted as a reference [112]. It can be observed that the performance of the WR semi-blind is better than that of the conventional LMMSE based estimator. Also, it can be found that the joint DFT and WR based channel estimation provides enhanced performance compared to WR, and its performance is closer to the CRB

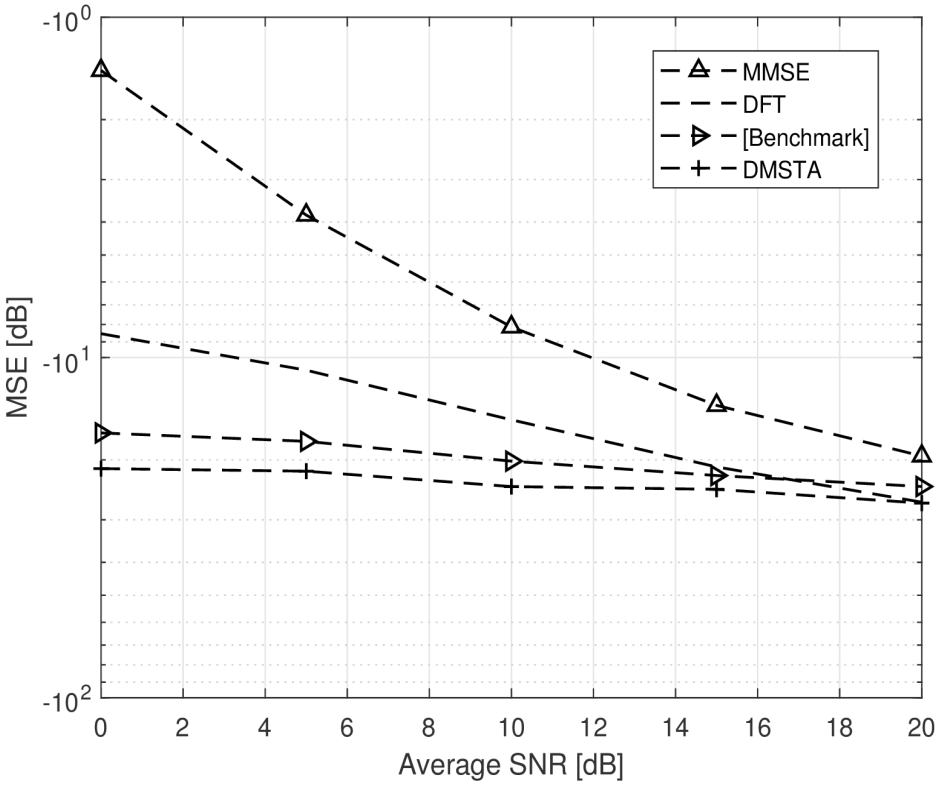

Figure 7-5: MSE of the conventional LMMSE, DFT and DMSTA and the modified DFT-based in [84] versus SNR.

than the conventional WR algorithm. Also, it can be seen that the estimation accuracy of the joint DFT-based and WR approaches can progressively improve toward the CRB with decreasing values of number of the paths.

Fig. 7-9 shows the uplink achievable rate for LMMSE with no pilot contamination, joint DFT and WR with L=100 under the effect of pilot contamination and the perfect channel estimation scenarios. It can be seen that the performance of the proposed estimator of the joint DFT and WR has closed performance to the perfect channel estimation and better performance than the LMMSE with no pilot contamination which demonstrate a significant improvement in estimation accuracy and addressing the pilot contamination problem.

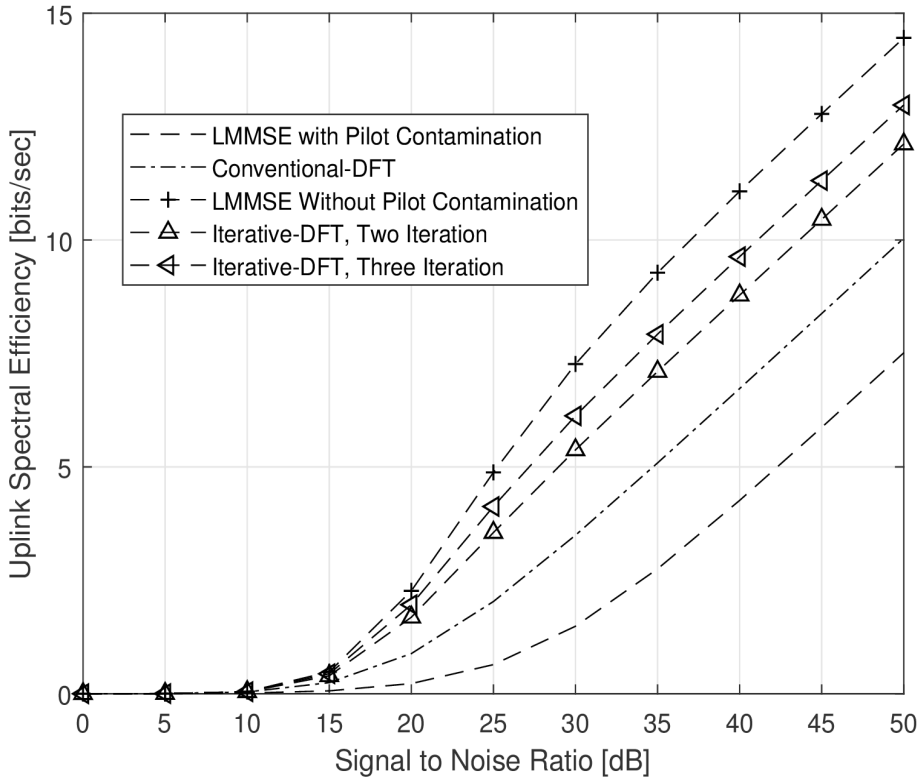

Figure 7-6: Uplink achievable rate for LMMSE with no pilot contamination, conventional DFT, Iterative DFT and LMMSE under the effect of strong pilot contamination versus SNR.

7.7 Conclusion

In this chapter, we propose three modified DFT-based channel estimation techniques. In the first part of the chapter, the proposed estimators tackled the pilot contamination problem by applying the iterative DFT and the MST approaches. The simulation results showed that the proposed technique provides much better performance compared to the conventional methods in terms of addressing the pilot contamination problem. Furthermore, the estimation performance of the DFT-based based estimator can be enhanced by increasing the number of subcarriers. In the last part of the chapter, We enhanced the DFT estimation by combining DFT with WR. Simulation results demonstrated improved estimation

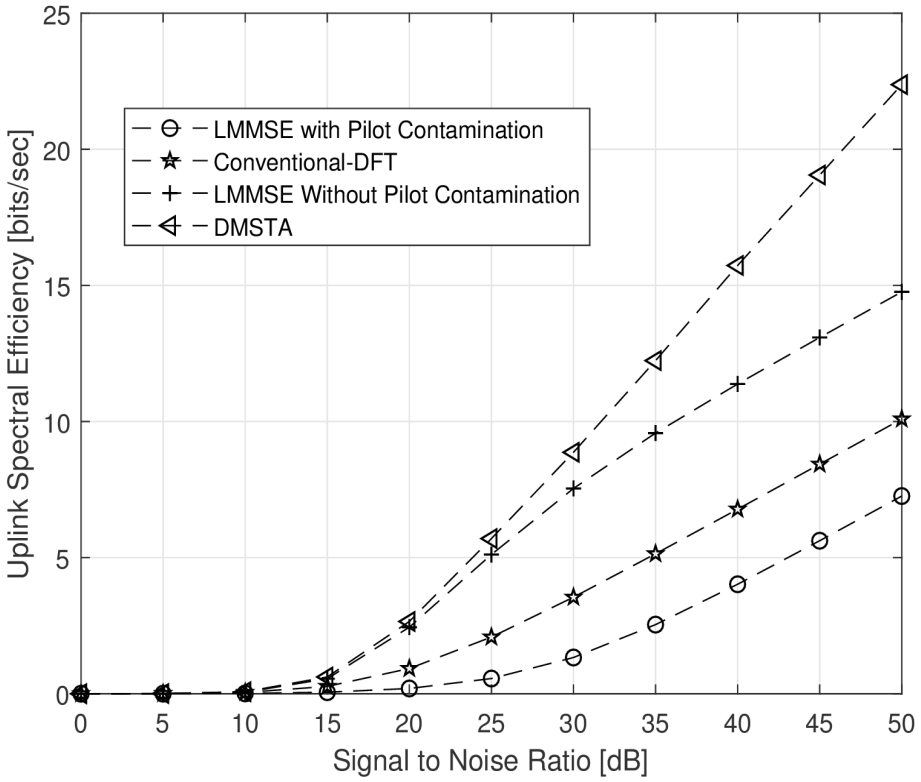

Figure 7-7: Uplink achievable rate for LMMSE with no pilot contamination, conventional DFT, DMSTA and LMMSE under the effect of strong pilot contamination versus SNR.

accuracy performance and significant addressing to the pilot contamination problem of the combined technique compared to the classical WR, the LMMSE and the reference CRB. In addition, it can be concluded that the estimation accuracy of the combined techniques can also be improved by decreasing the values of the number of the paths.

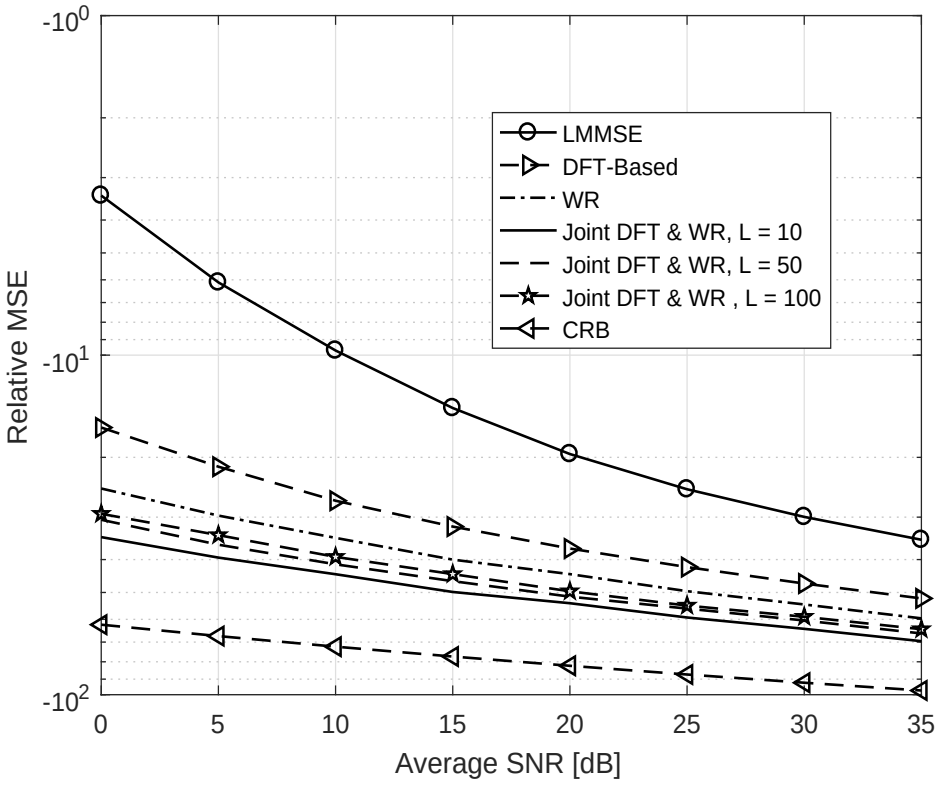

Figure 7-8: Relative MSE versus SNR of the conventional LMMSE, theoretical and practical WR, DFT, the joint DFT and WR based channel estimation for L= 10, 50 and 100 and CRB bound channel estimation for massive MIMO system

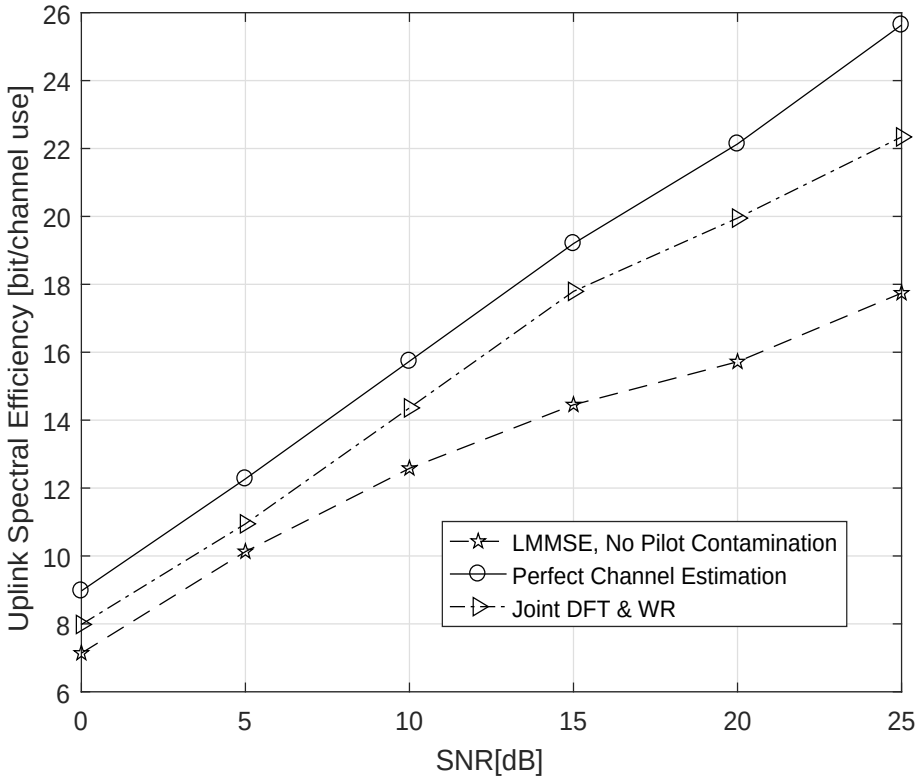

Figure 7-9: Achievable Uplink Rate versus SNR for the conventional LMMSE, the Joint DFT and WR based channel estimation techniques with Pilot Contamination and the Perfect Channel Estimation.

Chapter 8

Conclusions and Future Works

In this chapter, the main conclusions of the works in this thesis are presented. Some interesting future research directions are also explored in Section 8.2.

8.1 Summary of Results and Contributions

This thesis dealt with the pilot contamination reduction/elimination techniques for multi-user multi-cell massive multiple input multiple output (MIMO) systems. Several methods have been proposed based on different approaches i.e. Bayesian learning, semi-blind channel estimation, time domain channel estimation, optimal pilot design and the performance of pilot contamination under the effect of Rician and Nakagami-m fading channel models. The main contributions of the thesis are summarised below with concluding remarks.

In Chapter 3, we proposed the Bayesian compressed sensing (BCS) technique as a channel estimation algorithm to overcome the compressed sensing (CS) limitation. The BCS exploits the prior knowledge of the channel sparsity for better estimation accuracy. The pilot contamination problem addressed by employing fewer number of pilots than required by the conventional method. We proposed an enhancement approach to the performance of the BCS-based estimator via the principle of thresholding, which selected the most significant taps. The multi

task-BCS (MT-BCS) was also proposed to exploit the common statistical sparsity distribution to enhance the estimation accuracy performance. The Cramer Rao bound (CRB) has been derived as a reference line. The results showed that the BCS-based channel estimation algorithm could address the pilot contamination problem and provide better estimation accuracy compared to conventional methods.

In Chapter 4, the sparse Bayesian learning (SBL)-based channel estimation was proposed for a multi-cell massive MIMO system based on a pattern coupled hierarchical Gaussian framework. In the proposed technique, the sparsity of each channel coefficient is controlled by its own hyperparameter and the hyperparameters of its immediate neighbours. Closed form expression was derived for the mean square error (MSE) for the proposed technique. The approximated expression for the optimal pilot was derived to improve the estimation accuracy of the proposed algorithm using the Lagrange multiplier optimisation method. The CRB has been derived as a reference line. The results demonstrated the effectiveness of the proposed techniques to address pilot contamination in terms of MSE.

In Chapter 5, we addressed the pilot contamination problem in multi-cell massive MIMO systems by proposing a novel optimal pilot design to minimise the total MSE of the minimum mean square error (MMSE) estimators of all base stations subject to a transmit power constraint for each user in the network. By decomposing the non-convex MSE expression into distributed sub-problems at the base stations, each base station optimises its own pilot signals given the knowledge of pilot signals from the neighbouring base stations.Then, a successive optimisation is performed to transform each optimisation problem into convex form to be solved by the available optimisation packages. The results indicated that the proposed approach could reduce/mitigate pilot contamination efficiently.

In Chapter 6, we investigated the multi-cell massive MIMO over correlated Rician and correlated Nakagami-m fading channel models. We showed that the

estimation accuracy could be enhanced by increasing the values of the Rician fading factor by increasing the line-of-sight (LOS) component. Moreover, the achievable uplink rate and the MSE tools demonstrated that the pilot contamination problem could be addressed by increasing the values of the m-shaping factor and the K-Rician factor.

In Chapter 7, we investigated the discrete Fourier transform (DFT) channel estimation techniques for massive MIMO. The channel frequency response (CFR) was first estimated by applying the MMSE estimation, then it will transferred to the time domain by applying the inverse DFT (IDFT) to obtain the channel impulse response (CIR), a truncation of the CIR to remove the noise and interference is then performed. Thereafter, the CIR is transferred to CFR by applying the DFT. In addition, we proposed two modified DFT estimation algorithms by applying the same DFT procedure several times to obtain better estimation accuracy or by applying a threshold to the CIR to obtain the most significant taps. Further enhancement for DFT estimator has been proposed by proposing the joint DFT and whiting rotation (WR) channel estimation.

8.2 Suggestions for Future Research

With regard to future work, we identify a few potential problems, as follows.

8.2.1 Temporal Correlation-SBL

In chapter 4, we exploited the spatial correlation of the channel. However, the wireless channel is also subject to temporal correlation. Further work must conducted to exploit the temporal and spatial correlation to enhance the estimation accuracy.

8.2.2 Frequency Division Duplex (FDD) Mode for Massive MIMO systems

As explained in Chapter 1, massive MIMO can work in TDD and FDD modes, in FDD, the channel is estimated by users using a downlink training sent by BS. The estimated channel is then fed back to the BS. Thus, the pilot overhead is proportional to the number of antennae, which makes FDD impractical when the antenna array at the base station becomes large. However, under some assumptions, it is possible to use FDD to learn the channels in the large antenna array case. One possible solution to reduce the pilot overhead is to employ the CS-based estimator. However, as we explained in previous chapters, the CS estimation algorithms suffer from practical limitations. To address the CS limitation, we proposed SBL methods for FDD mode as a research challenge and as a work extension of this research.

8.2.3 Superimposed based channel estimation

In superimposed channel estimation, the pilot will be superimposed with data in such a way that the pilot contamination problem will be cancelled. By definition, the superimposed is not a channel estimation, but an arrangement of the pilot with data, thus it would be good to combine it with other estimation algorithms and to modify it in such a way as to reduce the energy efficiency.

8.2.4 Filter bank for Massive MIMO systems

The filter bank is a candidate technology for 5G that has good features when combined with massive MIMO, as the blind channel estimation of the filter bank will cancel the pilot contamination problem. Thus, further enhancement must be conducted to overcome blind estimation shortages.

8.2.5 Millimetre-Wave and small cells

In addition to massive MIMO, the millimetre wave and small cells are candidates for 5G. The small cells and millimeter wave can address the pilot contamination problem [113] and [3]. The study of pilot contamination using the combination of these technologies will be an important research direction in the future.

8.2.6 Hardware Impairment

Massive MIMO uses low-cost components, which may lead to hardware imperfections and cause distortion of the signal. In all the chapters, an ideal hardware is assumed. The pilot contamination issue in the non-ideal case is actually easier to mitigate compared to the ideal case [4], [67]. It would be interesting to include the impact of hardware impairment as extension of the proposed methods.

Appendix A

MMSE channel estimate $\widehat{\mathbf{H}}_{c^*,c^*}$ approximation

Reference (5.9), the term $E\{\mathbf{H}_{c^*,c^*}(\mathbf{Y}_{c^*})^H\}$ can be obtained as follow

$$E\{\mathbf{H}_{c^*,c^*}(\mathbf{Y}_{c^*})^H\} = E\{\mathbf{H}_{c^*,c^*}(\mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}^{\frac{1}{2}}_{c,c^*}\mathbf{H}_{c,c^*} + \mathbf{V}_{c^*})^H\}, \tag{A.1}$$

$$E\{\mathbf{H}_{c^*,c^*}(\mathbf{Y}_{c^*})^H\} = E\{\mathbf{H}_{c^*,c^*}\mathbf{H}^H_{c^*,c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{X}^H_{c^*}\}, \tag{A.2}$$

$$E\{\mathbf{H}_{c^*,c^*}(\mathbf{Y}_{c^*})^H\} = M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{X}^H_{c^*}. \tag{A.3}$$

While the term $\mathbf{\Omega}_{c^*} = E\{\mathbf{Y}_{c^*}(\mathbf{Y}_{c^*})^H\}$ can be obtained as follow

$$E\{\mathbf{Y}_{c^*}(\mathbf{Y}_{c^*})^H\} = E\{\mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}^{\frac{1}{2}}_{c,c^*}\mathbf{H}_{c,c^*} + \mathbf{V}_{c^*}(\mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}\mathbf{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}^{\frac{1}{2}}_{c,c^*}\mathbf{H}_{c,c^*} + \mathbf{V}_{c^*})^H\}, \tag{A.4}$$

$$E\{\mathbf{Y}_{c^*}(\mathbf{Y}_{c^*})^H\} = M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau. \tag{A.5}$$

$$\mathbf{\Omega}_{c^*} = M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau. \tag{A.6}$$

Now we can obtained the approximated $\widehat{\mathbf{H}}_{c^*,c^*}$ as in (5.10).

Appendix B

MSE closed form expression for the MMSE estimator

Reference (5.12) and (5.13) the mean square error can be written as

$$MSE_{c^*} = tr[E\{(\hat{\mathbf{H}}_{c^*,c^*} - \mathbf{H}_{c^*,c^*})(\hat{\mathbf{H}}_{c^*,c^*} - \mathbf{H}_{c^*,c^*})^H\}] \tag{B.1}$$

$$MSE_{MMSE} = tr[E\{\hat{\mathbf{H}}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H - \hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H - \mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H + \mathbf{H}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\}] \tag{B.2}$$

The term $E\{\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\}$ can be simplified as follow

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\} = E\{M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} \mathbf{Y}_{c^*}(\mathbf{H}_{c^*,c^*})^H\} \tag{B.3}$$

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\} = E\{M\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} (\mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{H}_{c^*,c^*} + \sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}^{\frac{1}{2}}\mathbf{H}_{c^*,c} + \mathbf{V}_{c^*})(\mathbf{H}_{c^*,c^*})^H\}, \quad \text{(B.4)}$$

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\} = E\{M\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} (\mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\mathbf{H}_{c^*,c^*})(\mathbf{H}_{c^*,c^*})^H\}, \quad \text{(B.5)}$$

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\} = E\{M\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} (\mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}})M\}, \quad \text{(B.6)}$$

since $E\{\mathbf{H}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\} = M\mathbf{I}_N$.

While the term $E\{\mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\}$ can be simplified mathematically as follow

$$E\{\mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = E\{\mathbf{H}_{c^*,c^*}(\mathbf{Y}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1}\mathbf{X}_{c^*} \mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\}M \quad \text{(B.7)}$$

$$E\{\mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = E\{(\mathbf{H}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}(\mathbf{X}_{c^*})^H + \mathbf{H}_{c^*,c^*}(\textstyle\sum_{c=1,c\neq c^*}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}^{\frac{1}{2}}\mathbf{H}_{c^*,c})^H + \mathbf{V}_{c^*}^H)[M\textstyle\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1}\mathbf{X}_{c^*}\mathbf{D}_{c^*,c^*}^{\frac{1}{2}}\}M \quad \text{(B.8)}$$

$$E\{\mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = M(\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} \mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}M \quad \text{(B.9)}$$

Now the term $E\{\hat{\mathbf{H}}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\}$ can be written as follow

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = E\{M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} \mathbf{Y}_{c^*}(\mathbf{Y}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}]^{-1}\mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}M\} \quad \text{(B.10)}$$

$$E\{\mathbf{Y}_{c^*}(\mathbf{Y}_{c^*})^H\} = [M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau], \quad \text{(B.11)}$$

$$E\{\hat{\mathbf{H}}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + \sigma^2\mathbf{I}_\tau]^{-1} \mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}M \quad \text{(B.12)}$$

It can be seen that $E\{\hat{\mathbf{H}}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\} = E\{\mathbf{H}_{c^*,c^*}(\hat{\mathbf{H}}_{c^*,c^*})^H\}$ So,

$$MSE_{c^*} = tr[E\{-\hat{\mathbf{H}}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H + \mathbf{H}_{c^*,c^*}(\mathbf{H}_{c^*,c^*})^H\}] \quad \text{(B.13)}$$

So, the closed form expression can be written as

$$MSE_{c^*} = tr[\{M\mathbf{I}_N - M\mathbf{D}^{\frac{1}{2}}_{c^*,c^*}(\mathbf{X}_{c^*})^H[M\sum_{c=1}^{C}\mathbf{X}_c\mathbf{D}_{c^*,c}(\mathbf{X}_c)^H + M\sigma^2\mathbf{I}_\tau]^{-1} (M\mathbf{X}_{c^*}\mathbf{D}^{\frac{1}{2}}_{c^*,c^*})\}] \quad \text{(B.14)}$$

Bibliography

[1] Nokia Solutions and Networks, "Technology Vision 2020 White Paper," pp. 44–51, May 2014.

[2] V. Jungnickel and K. Manolakis and W. Zirwas and B. Panzner and V. Braun and M. Lossow and M. Sternad and R. Apelfrojd and T. Svensson, "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, May 2014.

[3] E. Larsson et al., "Massive MIMO for Next Generation Wireless Systems", IEEE Communication Magazine, vol. 52, no. 2, pp. 186-95, Feb. 2014.

[4] V. L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, R. Zhang, "An overview of massive MIMO: Benefits and challenges", IEEE Journal on Selected Topics in Signal Processing, vol. 8, no. 5, pp. 742-758, Oct. 2014.

[5] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO-5G system", IEEE Communication Surveys Tuts., vol. 18, no. 2, pp. 905-923, 2nd Quart. 2016.

[6] E. Bjornson, E. G. Larsson, T. L. Marzetta, "Massive MIMO: ten myths and one critical question", IEEE Communication Magazine, vol. 54, no. 2, pp. 114-123, Feb. 2016.

[7] M. Masood, L. Afify, and Tareq Y. Al-Naffouri, "Efficient Coordinated Recovery of Sparse Channels in Massive MIMO," IEEE Transactions on Signal Processing, vol. 63, no. 1, Jan. 2015.

[8] J. Zhang, B. Zhang, S. Chen et al., "Pilot contamination elimination for large-scale multiple-antenna aided OFDM systems", IEEE Journal of Selected Topics in Signal Processing", pp. 759-772, March 2014.

[9] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Transaction Wireless Communication*, Nov. 2010, vol.9, no. 11, pp. 3590-3600.

[10] Z. Gong, C. Li, F. Jiang, " Pilot contamination mitigation strategies in massive MIMO systems," *IET Communications*, Nov. 2017, Vol. 11, no. 16, pp. 2403-2409.

[11] M. Sadeghi, E. Bjornson, E. G. Larsson, C. Yuen, and T. L. Marzetta, "Max-min fair transmit precoding for multi-group multicasting in massive MIMO," *IEEE Transaction on Wireless Communication*, Dec. 2017, Vol. 17, pp. 1536-1276.

[12] L. Zhao, G. Geraci, T. Yang, et al.:"A Tone-Based AoA Estimation and Multiuser Precoding for Millimeter Wave Massive MIMO," *IEEE Transactions on Communications*, Vol. 65, Dec. 2017, pp. 5209-5225.

[13] S. Zarei, J. Aulin, W. Gerstacker and R. Schober: "Max-Min Multicell-Aware Precoding and Power Allocation for Downlink Massive MIMO Systems," *IEEE Signal Processing Letters*, Vol. 24, Oct. 2017, pp. 1433-1437.

[14] N.Moghadam, H. Shokri-Ghadikolaei, G. Fodor, M. Bengtsson and C. Fischione: "Pilot Precoding and Combining in Multiuser MIMO Networks," *IEEE Journal on Selected Areas in Communications*, Vol. 35, July 2017, pp. 1632-1648.

[15] H. H. Yang, G. Geraci, T. Q. S. Quek and J. G. Andrews:" Cell-Edge-Aware Precoding for Downlink Massive MIMO Cellular Networks, " *IEEE Transactions on Signal Processing*, Vol. 65, July 2017, pp. 3344-3358.

[16] Fernandes, F., Ashikhmin, A., Marzetta, T.L.: "Inter-cell interference in noncooperative TDD large scale antenna systems", IEEE Journal Selected Areas of Communication, 2013, 31, (2), pp. 192-201

[17] J. Jose, A. Ashikhmin, T. L. Marzetta, S. Vishwanath, "Pilot contamination and precoding in multi-cell TDD systems", IEEE Transaction on Wireless Communications, vol. 10, no. 8, pp. 2640-2651, Aug. 2011.

[18] Jose, J., Ashikhmin, A., Marzetta, T.L., et al.: "Pilot contamination problem in multi-cell TDD systems", Proc. IEEE International Symposium of Information Theory, Seoul, pp. 2184 -2188, June 2009.

[19] Peng Xu, Jinkuan Wang and Jiangzhou Wang, "Effect of pilot contamination on channel estimation in massive MIMO systems", 2013 International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1-6, Oct. 2013.

[20] H. Zhang, S. Gao, D. Li, H. Chen and L. Yang, "On Superimposed Pilot for Channel Estimation in Multi-cell Multiuser MIMO Uplink: Large System Analysis", IEEE Transaction on Vehicular Technology, pp. 1492 - 1505, vol. 65, no. 3, March 2015.

[21] C. Qi, Y. Huang, S. Jin, L. Wu, "Sparse channel estimation based on compressed sensing for massive MIMO systems", Proceeding of IEEE International. Conference on Communications (ICC), pp. 4558-4563, Jun. 2015.

[22] S. L. H. Nguyen, A. Ghrayeb, "Compressive sensing-based channel estimation for massive multiuser MIMO systems", Proc. IEEE Wireless Communication Network Conference, pp. 2890-2895, Apr. 2013.

[23] Y. Nan, L. Zhang, X. Sun, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for TDD massive MU-MIMO systems", IEEE Wireless Communication Letter, vol. 4, no. 4, pp. 345-348, Aug. 2015.

[24] D. Verenzuela, E. Bjoernson and L. Sanguinetti "Joint UL and DL Spectral Efficiency Optimization of Superimposed Pilots in Massive MIMO," *Globecom Workshops (GC Wkshps)*, 4-8 Dec. 2017.

[25] D. Verenzuela, E. Bjoernson, L. Sanguinetti, "Spectral Efficiency of Superimposed Pilots in Uplink Massive MIMO Systems"*GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 4-8 Dec. 2017.

[26] K. Upadhya, S. A. Vorobyov, M. Vehkapera, "Superimposed Pilots Are Superior for Mitigating Pilot Contamination in Massive MIMO," *IEEE Transactions on Signal Processing*, Vol. 65, June 2017, pp. 2917-2932

[27] J. Ma, C. Liang, C. Xu, L. Ping, "On Orthogonal and Superimposed Pilot Schemes in Massive MIMO NOMA Systems," *IEEE Journal on Selected Areas in Communications*, Vol. 35, Dec. 2017, pp. 2696-2707

[28] Y. Chen, "Low-Cost Superimposed Pilots Based Receiver for Massive MIMO in Multicarrier System," *IEEE Vehicular Technology Conference (VTC Spring)*, 4-7 June 2017.

[29] F. Li, H. Wang, M. Ying, W. Zhang,"Channel estimation using superimposed pilots and second-order statistics for massive MIMO networks," *IEEE 18th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 3-6 July 2017.

[30] H. Zhang, S. Gao, D. Li, H. Chen and L. Yang, "On Superimposed Pilot for Channel Estimation in Multi-cell Multiuser MIMO Uplink: Large System

Analysis", *IEEE Transaction on Vehicular Technology*, March 2015, **65**, no. 3, pp. 1492 - 1505.

[31] S. Nguyen, "Compressive Sensing for Multi-channel and Large-scale MIMO Networks",PhD. thesis, Department of Electrical And Computer Engineering, Concordia University, Montreal, Quebec, Canada, 13 Jan 2014.

[32] W. U. Bajwa, J. Haupt, A. M. Sayeed, R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels", Proceeding IEEE, vol. 98, no. 6, pp. 1058-1076, June 2010.

[33] R. G. Baraniuk, "Compressive sampling", IEEE Signal Process. Magazine, vol. 24, no. 4, pp. 118-124, July 2007.

[34] M. Carlin, P. Rocca, G. Oliveri, F. Viani, and A.Massa,"Directions-of-arrival estimation through Bayesian compressive sensing strategies",IEEE Transaction Antennas Propagation, vol.61, no. 7, pp.3828-3838, July 2013.

[35] R. Muller, M. Vehkapera, and L. Cottatellucci, "Blind pilot decontamination", in Proceeding ITG Workshop Smart Antennas , Stuttgart, Germany, Mar. 2013, pp. 1-6.

[36] R. Muller, L. Cottatellucci, and M. Vehkapera, "Blind pilot decontamination", IEEE Journal on Selected Topics on Signal Processing, vol. 8, no. 5, pp. 773-786, Oct. 2014.

[37] A. Chockalingam and B. Sundar Rajan, "Large MIMO Systems," 1st edition: Cambridge: Cambridge University Press, 2014

[38] Jerry R. Hampton, "Introduction to MIMO Communications," 1st edition: Cambridge: Cambridge University Press, 2014.

[39] Yong Soo Cho, Jaekwon Kim, Won Young Yang and Chung G. Kang, "MIMO-OFDM Wireless Communications with MATLAB", John Wiley & Sons (Asia) Ltd, 2010.

[40] A. Hamza, S. Khalifa, H. Hamza, and K. Elsayed, "A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks," IEEE Communication Surveys Tuts., vol. 15, no. 4, pp. 1642–1670, 4th Quart. 2013.

[41] S. R. Saunders, Antenna And Propagation For Wireless Communication Systems. New York: John Wiley & Sons, 2000.

[42] J. G. Proakis, Digital Communications, 4th edition New York: McGraw-Hill, 2001.

[43] Aditya K. Jagannatham, "Bandwidth Efficient Channel Estimation for Multiple-Input Multiple-Output (MIMO) Wireless Communication Systems: A Study of semi-Blind and Superimposed Schemes", Ph.D. dissertation, Department of ECE, University of California San Diego, California, USA, 2007.

[44] M. K. Simon and M. S. Alouini, Digital Communication over Fading Channels, 2nd ed. John Wiley Sons, 2004.

[45] Z. Kang, K. Yao, and F. Lorenzelli, "Nakagami-m fading modeling in the frequency domain for OFDM system-performance analysis," IEEE Communications Letters, vol. 7, pp. 484–486, October 2003.

[46] M. Nakagami, The m distribution-A general formula of intensity distribution of rapid fading; Statistical Methods in RadioWave Propagation. W. G. Hoffman, Ed. Oxford, UK.: Pergamon, 1960.

[47] Y. Liu, Z. Tan, H. Hu, L. Cimini, G. Li, "Channel estimation for OFDM", IEEE Communication Surveys and Tutorial, vol. 16, no. 4, pp. 1891-1908, 2014.

[48] Simon Haykin and Michael Moher, "Modern Wireless Communications". USA, Prentice Hall, 2005.

[49] Thomas A. Schonhoff, Arthur Anthony Giordano, "Detection and Estimation Theory and Its Applications Theory Practice and Matlab". USA, Prentice Hall, 2006.

[50] Y. C. Eldar and G. Kutyniok, Compressed Sensing Theory and Applications. Cambridge University Press, 2012.

[51] A. Scherb, K. Kammeyer, "Bayesian channel estimation for doubly correlated MIMO systems", Proceeding of IEEE Workshop Smart Antennas, 2007.

[52] S. M. Kay. Fundamentals of Statistical Signal Processing: Estimation Theory. PTR Prentice-Hall, Englewood cliffs, New Jersey, 1993.

[53] Sina Jafarpour, "Deterministic Compressed Sensing", PhD. thesis, Department of Computer Science, Princeton University, Princeton, New Jersey, USA, November 2011.

[54] Depeng Yang, "Turbo Bayesian Compressed Sensing", PhD thesis, University of Tennessee, Knoxville, August 2011.

[55] Z.-Q. Luo, W. Yu, "An introduction to convex optimization for communications and signal processing", IEEE Journal of Selected Areas on Communications, vol. 24, no. 6, pp. 1426-1438, Aug. 2006.

[56] H. Hindi, "A Tutorial on Convex Optimization", Proceeding of American Control Conference, Boston, USA, 2004.

[57] L. Vandenberghe, S. Boyd, K. Comanor, "Generalized Chebyshev bounds via semidefinite programming", SIAM Review 49, 52-64, 2007

[58] Xinying Guo, Sheng Chen, Jiankang Zhang, Xiaomin Mu, Lajos Hanzo,"Optimal Pilot Design for Pilot Contamination Elimination/Reduction

in Large-Scale Multiple-Antenna Aided OFDM Systems", IEEE Transaction on Wireless Communications, Vol. 15, No.11, Nov. 2016.

[59] M. R. Nakhai, "Multicarrier transmission", IET Signal Processing, vol. 2, no. 1, pp. 114, Mar. 2008.

[60] C. Qi, G. Yue, L. Wu, Y. Huang and A. Nallanathan, "Pilot design schemes for sparse channel estimation in OFDM systems", IEEE Transaction of Vehicular Technology, vol. 64, no. 4, pp. 1493-1505, Apr. 2015.

[61] M. Masood, L. Afify, and Tareq Y. Al-Naffouri, "Effcient Coordinated Recovery of Sparse Channels in Massive MIMO," IEEE Transactions on Signal Processing, vol. 63, no. 1, Jan. 2015.

[62] W. Ding, F. Yang, W. Dai and J. Song, "Time-frequency joint sparse channel estimation for MIMO-OFDM systems", IEEE Communication Letter, vol. 19, no. 1, pp. 58-61, Jan. 2015.

[63] S. Ji, Y. Xue and L. Carin, "Bayesian compressive sensing", IEEE Transaction on Signal Processing, vol. 56, no. 6, pp. 2346-2356, June 2008.

[64] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine", Journal of Machine Learning Research, vol. 1, pp. 211-244, 2001.

[65] X. Cheng, M. Wang, and Y. Guan, "Ultra wideband Channel Estimation: A Bayesian Compressive Sensing Strategy Based on Statistical Sparsity", IEEE Transaction on Vehicular Technology, Vol. 64, no. 5,pp. 1819 - 1832, May 2015.

[66] S. Ji, D. Dunson and L. Carin, "Multi-task compressive sensing", IEEE Transaction on Signal Processing, vol. 57, no. 1, pp. 92-106, Jan. 2009.

[67] A. K. Jagannatham and B. D. Rao, "Whitening-Rotation-Based Semi-Blind MIMO Channel Estimation", IEEE Transaction on Signal Processing, vol. 54, no. 3, pp. 861-869, 2006.

[68] J. T. Parker, P. Schniter and V. Cevher, "Bilinear generalized approximate message passing", IEEE Transaction on Signal Processing, vol. 62, no. 22, pp. 5839-5853, Nov. 2014.

[69] J. A. Tropp, A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit", IEEE Transaction of Information Theory, vol. 53, no. 12, pp. 4655-4666, Dec. 2007.

[70] Olutayo O. Oyerinde, "Reweighted regularised variable step size normalised least mean square-based iterative channel estimation for multicarrier-interleave division multiple access systems", IET Signal Processing, Volume: 10, no. 8, Oct. 2016.

[71] Hien Quoc Ngo et al.: "Analysis of the Pilot Contamination Effect in Very Large Multicell Multiuser MIMO Systems for Physical Channel Models", in Proceeding of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2011, Prague, Czech.

[72] Y. Shen, H. Duan, J. Fang, H. Li, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals", IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP), May 2014, pp. 1896-1900.

[73] J. Fang, Y. Shen, H. Li and P. Wang, "Pattern-Coupled Sparse Bayesian Learning for Recovery of Block-Sparse Signals", IEEE Transaction on Signal Processing, 2015, 63, no. 2, pp. 360-372.

[74] Y. Barbotin, M. Vetterli, "Estimation of sparse MIMO channels with common support", IEEE Transaction on Communications, Dec. 2012, 60, no. 12, pp. 3705-3716.

[75] Y. Nan, L. Zhang, X. Sun, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for TDD massive MU-MIMO systems", IEEE Wireless Communication Letter, Aug. 2015, 4, no. 4, pp. 345-348.

[76] Y. Barbotin, M. Vetterli, "Fast and robust parametric estimation of jointly sparse channels", IEEE J. Emerging Sel. Topics Circuits Syst., vol. 2, no. 3, pp. 402-412, Sep. 2012.

[77] M. Patzold and G. Rafiq, "Sparse multipath channels: modelling analysis, and simulation," 24th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications", PIMRC 2013. London, UK, Sep. 2013, pp. 30-35.

[78] I. Barhumi, G Leus, M Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels", IEEE Transaction on Signal Processing, June 2003, 51, pp. 1615-1624.

[79] M Biguesh, AB Gershman, "Downlink channel estimation in cellular systems with antenna arrays at base stations using channel probing with feedback", EURASIP Journal Applied Signal Processing, August 2004, no.9, pp. 1330-1339.

[80] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals," IEEE Transaction on Signal Processing, Mar. 2006,54, no. 3, pp. 884-893.

[81] K. B. Petersen and M. S. Pedersen, "The Matrix Cookbook", Technical University of Denmark, 2012.

[82] M. Ozgor, S. Erkucuk, H.A. Cirpan, "Bayesian compressive sensing for ultra-wideband channel estimation: algorithm and performance analysis", Telecommunication Systems, 2015, 59, pp. 417-427.

[83] Z. Ben-Haim, Y. C. Eldar, "A lower bound on the Bayesian MSE based on the optimal bias function", IEEE Transactions on Information Theory, 2009, 55, no. 11,pp. 5179-5196.

[84] S. M. Kay. "Fundamentals of Statistical Signal Processing: Estimation Theory", PTR Prentice-Hall, Englewood Cliffs, 1993, New Jersey.

[85] M. C. Grant and S. P. Boyd, The CVX Users′ Guide, Release 2.1., Mar. 2015, [Online]. Available: http://web.cvxr.com/cvx/doc/CVX.pdf.

[86] Ohno S, Manasseh E, Nakamoto M, "Preamble and pilot symbol design for channel estimation in OFDM systems with null subcarriers", EURASIP Journal on Wireless Communications and Networking,pp.1-17, Dec. 2011.

[87] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge University Press, 2004.

[88] X. Zhu et al., "Smart pilot assignment for Massive MIMO," IEEE Communication Letters, vol. 19, no. 9, pp. 1644−1647, Sep. 2015.

[89] L. You et al., "Pilot reuse for Massive MIMO transmission over spatially correlated rayleigh fading channels," IEEE Transaction Wireless Communication, vol. 14, no. 6, pp. 3352−3366, Jun. 2015.

[90] Emil Bjornson, Jakob Hoydis, Marios Kountouris, Merouane Debbah, "Massive MIMO Systems with Non-Ideal Hardware: Energy Efficiency, Estimation, and Capacity Limits," IEEE Transactions on Information Theory, vol. 60, no. 11, pp. 7112−7139,November 2014.

[91] P. Ciblat, P. Bianchi, M. Ghogho, "Optimal Training for Frequency Offset Estimation in Correlated-Rice Frequency-Selective Channel", Signal Processing Advances in Wireless Communications, July 2008.

[92] S. Zhu, et al: "Probability Distribution of Rician K-Factor in Urban,Suburban and Rural Areas Using Real-World Captured Data," IEEE Transaction on Antennas and Propagation, vol. 62, no. 7, pp. 3835-3839, July 2014.

[93] Li Tang, Zhu Hongbo, "Analysis and Simulation of Nakagami Fading Channel with MATLAB", Asia- Pacific, CEEM′ 2003, Nov, 4-7, 2003, pp. 490-494 Hangzhou, China.

[94] Hamid Nooralizadeh et al.: "Appropriate Algorithms for Estimating Frequency-Selective Rician Fading MIMO Channels and Channel Rice Factor: Substantial Benefits of Rician Model and Estimator Tradeoffs," EURASIP Journal on Wireless Communications and Networking, vol. 2010, 17 August 2010.

[95] Emil Bjornson and Bjorn Ottersten, "A Framework for Training-Based Estimation in Arbitrarily Correlated Rician MIMO Channels with Rician Disturbance," IEEE Transactions on Signal Processing, vol. 58, no. 3, pp. 1807-1820, March 2010.

[96] Q. Zhang, S. Jin, K.-K. Wong, H. Zhu, and M. Matthaiou, "Power scaling of uplink massive MIMO systems with arbitrary -rank channel means," IEEE Journal on Selected Topics on Signal Processing, vol. 8, no. 5, pp. 966−981, Oct. 2014.

[97] S. L. Loyka "Channel capacity of MIMO architecture using the exponential correlation matrix", IEEE Communication Letter, vol. 5, no. 9, pp. 369 -371 2001.

[98] Further advancements for E-UTRA physical layer aspects (Release 9), 3GPP TS 36.814, Mar. 2010.

[99] Y. Kang, K. Kim and H. Park "Efficient DFT-based channel estimation for OFDM systems on multipath channels", IET Communications, 2007, 1, 2, pp. 197 -202.

[100] Y. Wang, L. Li, P. Zhang and Z. Liu "DFT-Based Channel Estimation with Symmetric Extension for OFDMA Systems", EURASIP Journal on Wireless Communication and Networking, 2009, 647130, 10.1155/2009/647130.

[101] C. Bai1, S. Zhang, S. Bai, et al.: "Development of discrete Fourier transform-based channel estimation algorithms for a coherent optical orthogonal frequency division multiplexing transmission system", IET Communications, 2014, 8, 14, pp. 2528-2534.

[102] Y. Qiao, S. Yu, P. Su, L. Zhang, "Research on an iterative algorithm of LS channel estimation in MIMO OFDM systems", IEEE Transaction on Broadcasting, vol. 51, no. 1, pp. 149-153, March 2005.

[103] Qiao et al. "Research on an iterative algorithm of LS channel estimation in MIMO OFDM systems", Proceedings of the IEEE 6th Circuits and Systems Symposium ,pp. 729-732, Piscataway, NJ, USA, IEEE, May 31, 2004.

[104] M. Jiang, G. Yue, N. Prasad et al., "Enhanced DFT-Based Channel Estimation for LTE Uplink", Proceeding of IEEE 75th Vehicular Technology Conference (VTC Spring), pp. 1-5, May 2012.

[105] Y.-S. Lee, H.-C. Shin, and H.-N. Kim, "Channel estimation based on a time-domain threshold for OFDM systems," IEEE Transactions on Broadcasting, vol. 55, no. 3, pp. 656−662, Sep. 2009.

[106] Y. Peng, G. C. Alexandropoulos and Y. Li "Pilot-assisted channel estimation with MSE-optimal thresholding for OFDM systems", Transaction on Emerging Telecommunication Technology, 23 May 2016.

[107] H. Wang; Z. Pan; J. Ni, et al.: "A Temporal Domain Based Method against Pilot Contamination for Multi-cell Massive MIMO Systems", IEEE 79th Vehicular Technology Conference (VTC Spring), Seoul, May 2014, pp 1-5.

[108] Aditya K. Jagannatham, "Bandwidth Efficient Channel Estimation for Multiple-Input Multiple-Output (MIMO) Wireless Communication Systems: A Study of semi-Blind and Superimposed Schemes", Ph.D. dissertation, Department of ECE, University of California San Diego, California, USA, 2007.

[109] F. Wan, W. P. Zhu and M. N. Swamy "A signal perturbation free whitening-rotation-based semiblind approach for MIMO channel estimation", IEEE Transaction Signal Processing, vol. 57, no. 8, pp.3154 -3166 2009

[110] Xia Liu, et al., "investigation into a Whitening-Rotation-Based Semi-blind MIMO Channel Estimation for Correlated Channels", IEEE 2nd International Conference on Signal Processing and Communication Systems (ICSPCS 08), December 2008, Gold Coast, Vol. 1, pp. 1-4.

[111] Q. Zhang et al.: "Whitening-Rotation-Based Semi-Blind Estimation of MIMO FIR Channels", International Conference on Wireless Communications and Signal Processing (WCSP), 13-15 Nov. 2009.

[112] N. K. Venkategowda and A. K. Jagannatham, "WR Based Semi-Blind Channel Estimation for Frequency-Selective MIMO MC-CDMA Systems," 2012 IEEE Wireless Communications and Networking Conference: PHY and Fundamentals, Paris, France, Apr. 2012, pp. 317-321.

[113] X. Ma, F. Yang, S. Liu, J. Song, Z. Han, "Design and optimization on training sequence for mmWave communications: A new approach for sparse channel estimation in massive MIMO", IEEE Journal on Selected Areas on Communications, vol. 35, no. 7, pp. 1486-1497, Jul. 2017.